Individualization of the Pubic Hair/Pubic Area Microbiome And Its Potential as an Indicator of Sexual Contact

A Dissertation
Presented to
The Academic Faculty

By

Diana W. Williams

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biology

Georgia Institute of Technology

May 2018

Individualization of the Pubic Hair/Pubic Area Microbiome
And Its Potential as an Indicator of Sexual Contact

Approved by:

Dr. Greg Gibson, Advisor
School of Biological Sciences
Georgia Institute of Technology

Dr. Annalise Paaby
School of Biological Sciences
Georgia Institute of Technology

Dr. Frank Stewart
School of Biological Sciences
Georgia Institute of Technology

Dr. Timothy Kalafut
USACIL
Defense Forensic Science Center

Dr. Joseph Lachance
School of Biological Sciences
Georgia Institute of Technology

Date Approved: March 27, 2018

To me, observing is much

better sport than writing.

Charles Darwin to Henry Fawcett

ACKNOWLEDGEMENTS

Thank you to everyone to has helped me on this journey. There are so many people who have played roles of support and encouragement. Thank you to the Defense Forensic Science Center (DFSC) for the opportunity to pursue this dream. My co-workers in the DNA unit had to take up the slack left behind by my temporary departure. I'm coming back soon! The DFSC Office of the Chief Scientist welcomed me with open arms and provided invaluable support. Dr. Rachel Creager was my most important sounding board and encouraged me to cite the RadioLab podcast as one of my sources. I couldn't have asked for a better Thesis Committee at Georgia Tech. They have been nothing but supportive throughout this journey. Dr. Greg Gibson took on my special circumstances and has supported me every step of the way. All of my friends and family have been so supportive and interested in my journey. Thank you to my parents for all of their unconditional love and support. Thank you to my children, Logan and Cate, for putting up with a more distracted mommy for the past few years. And the largest thanks go to my husband, Pat, for taking up the household slack when I had to study or work on my research. I couldn't have done it without him.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF TABLES viii

LIST OF FIGURES x

LIST OF SYMBOLS AND ABBREVIATIONS xiii

SUMMARY xvii

CHAPTER 1: INTRODUCTION 1

1.1 Investigation of Sexual Assaults 1

1.2 The Microbiome 3

1.3 Techniques 7

1.4 The Microbiome and Forensic Science 9

1.5 References 13

CHAPTER 2: INDIVIDUALIZATION OF PUBIC HAIR BACTERIAL COMMUNITIES AND THE EFFECTS OF STORAGE TIME AND TEMPERATURE 16

2.1 Abstract 16

2.2 Introduction 17

2.3 Methods 21

2.3.1 Sample Collection 21

2.3.2 DNA Extraction and Sequencing 22

2.3.3 Sequence Analysis and Community Comparison 24

2.4 Results 26

2.5 Discussion 35

2.6 Supplemental Figures and Tables 41

2.7 References 45

CHAPTER 3: CLASSIFICATION OF INDIVIDUALS AND THE POTENTIAL TO DETECT SEXUAL ACTIVITY USING THE PUBIC MICROBIOME 49

3.1 Abstract 49

3.2 Introduction 50

3.3 Methods 54

3.3.1 Sample Collection 54

3.3.2 DNA Extraction and Sequencing 56

3.3.3 Sequence Analysis and Community Comparisons 57

3.3.4 Deblur Sample Processing 59

3.3.5 Supervised Learning Classification 59

3.3.6 Hierarchical Clustering 61

3.3.7 Closest Non-self Neighbor Comparison 62

3.3.8 Comparison to Other Skin Sites and Buccal Mucosa 62

3.4 Results 63

3.4.1 Demographic Summary 63

3.4.2 Sequencing Summary 65

3.4.3 Alpha and Beta Diversity 67

3.4.4 Random Forests Modeling 72

3.4.5 Hierarchical Clustering 76

3.4.6 Closest Non-self Neighbor 80

3.5 Discussion 82

3.6 Supplemental Figures and Tables 90

3.7 References 97

CHAPTER 4: MIXTURES 102

4.1 Abstract 102

4.2 Introduction 103

4.3 Materials and Methods 107

4.4 Results 113

4.4.1 Shifts in Beta Diversity 113

4.4.2 Percent Mixing Required to Infer Transfer 114

4.4.3 Suspect Identification from the General Population 118

4.4.4 Culprit Identification from a Small Pool of Suspects 125

4.4.5 Exculpation of Falsely Accused from a Small Pool of Suspects 130

4.5 Discussion 131

4.6 References 139

CHAPTER 5: THE PATH FORWARD 140

5.1 The Pubic Area Microbiome and Forensics 140

5.2 Processing Low Biomass Samples and Stochastic Effects 142

5.3 Contamination 146

5.4 Mixture Analysis 149

5.5 Additional Considerations 153

5.6 Conclusion 154

5.2 References 155

LIST OF TABLES

Table 2.1	Effect sizes (R^2) calculated using the weighted UniFrac distance matrices and the vegan::adonis method	28
Table 2.2	OTUs at order level which are significantly differentially abundant using DESeq2	30
Table 2.3	Prediction of categories using Random Forests supervised classification within categories using OTUs as predictors	33
Supplemental Table 2.1	Read counts for OTUs at the gender level which are differentially abundant across genders	42
Supplemental Table 2.2	Comparison of similarity between weighted UniFrac distances within an individual's samples and another individual's samples using a two-sided Student's two-sample t-test	43
Supplemental Table 2.3	Random Forests (10-fold cross validation, 500 trees) label probabilities for each sample for the category individual	44
Table 3.1	Summarized demographic and survey responses of participants	64
Table 3.2	Effect sizes of various metadata categories with the vegan::adonis method based on weighted and unweighted UniFrac distances (hairs and swabs together)	71
Table 3.3	Random Forest 10-fold cross-validation estimated error based on 500 trees	72
Table 3.4	Predicted individual classification of storage study samples	75
Table 3.5	Comparison of Chi Square (couple by cluster) of actual couple assignment and random permutations	78
Supplemental Table 3.1	Metadata of HMP buccal mucosal samples	93

Supplemental Table 3.2 Pubic area summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances 94

Supplemental Table 3.3 Elbow summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances 95

Supplemental Table 3.4 Ear summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances 96

Table 4.1 Maximum, minimum, mean, and standard deviation for (a) Q1 percent OTU sharing, (b) weighted UniFrac Distance, and (c) Deblur percent OTU sharing for *in silico* mixtures 117

Table 4.2 ROC summary for Q1 and Deblur percent shared OTUs 128

Table 4.3 Means, maximums, minimums, and standard deviations of percentage of OTUs shared in population samples 130

LIST OF FIGURES

Figure 2.1	Taxonomic distribution at the genus level across storage time and temperature conditions	28
Figure 2.2	Effect of individual, gender, storage time, and storage temperature on the bacterial communities of the pubic hair	29
Figure 2.3	Comparison of weighted UniFrac distances within and between individuals	32
Figure 2.4	Comparison of Simpson Index of alpha diversity for each individual across storage time for each storage temperature	35
Supplemental Figure 2.1	Taxonomic distribution at the genus level without rarefaction across storage time and temperature conditions. Sorted by individuals	41
Figure 3.1	Genus level taxonomic bar chart of pubic area microbiome samples	66
Figure 3.2	Comparisons of Faith's Phylogenetic Diversity (y-axis) for various metadata categories	68
Figure 3.3	Principal components plots of the weighted (a-c) and unweighted (d-f) UniFrac distances for 'Individual' (a, d), 'Couple' (b, e), and Gender (c, f)	70
Figure 3.4	Grouping of couples in the hierarchical clustering when collapsed by individual and time point	77
Figure 3.5	Comparison of average number of sexual encounters in week prior to collection and percentage of hair/swab clusters shared at each time point for each couple	79
Figure 3.6	Proportion of closest-neighbor samples using Weighted and Unweighted UniFrac distances for ear, elbow, and pubic hair/swab samples for (a) self, partner, and others, (b) just partner or other, (c) and just partner or other for just the couples	81

Supplemental Figure 3.1 Contents of survey questions completed by participants at each collection time point 90

Supplemental Figure 3.2 Principal components analysis showing the relationship among microbiomes of various skin sites from Grice et al. based on taxon relative abundances 91

Supplemental Figure 3.3 Chi-square distributions of hair and swabs clusters for 8-15 clusters 91

Supplemental Figure 3.4 Chi-square distributions of swabs-only clusters for 8-15 clusters 92

Supplemental Figure 3.5 Chi-square distributions of hair-only clusters for 8-15 clusters 92

Figure 4.1 Weighted UniFrac PCoA plot of selected *in silico* couple mixtures 113

Figure 4.2 Density plots of various categorical groupings of *in silico* mixtures by (a) percentage of Q1 OTUs shared, (b) weighted UniFrac distance, and (c) percentage of Deblur OTUs Shared 116

Figure 4.3 Predicted sources to the pseudo-Couple 1 mixtures using the individuals in pseudo-Couple 1 as the source Environments 120

Figure 4.4 Predicted sources to the pseudo-Couple 1 mixtures using one individual in pseudo-Couple 1 and a non-contributing individual as source environments 121

Figure 4.5 Predicted sources to the pseudo-Couple 1 mixtures using all individuals as source environments 122

Figure 4.6 Variation in predicted contributor proportions in population samples as determined by SourceTracker 124

Figure 4.7 Percent OTU sharing ROC curves by contribution level for (a) Q1 OTUs and (b) Deblur OTUs 126

Figure 4.8 Precision and negative predictive value (NPV) of Q1 OTU sharing at 10% contribution level and 65% OTU sharing for equal prior likelihood and 10% prior likelihood 127

Figure 4.9 Density plot of percentage of OTUs shared for the categorical groupings of population samples using (a) Q1 generated OTUs and (b) Deblur generated OTUs 129

Figure 5.1 Filtered read counts as a function of input sample library concentration 144

Figure 5.2 Filtered read counts as a function of input sample library concentration with >2 million read outliers at <50 pM removed 145

Figure 5.3 PCoA plots of the weighted UniFrac distances at (a) 4428, (b) 1000, and (c) 500 read rarefaction 148

LIST OF SYMBOLS AND ABBREVIATIONS

~	Approximately
©	Copyright
>	Greater than
≥	Greater than or equal to
<	Less than
≤	Less than or equal to
μl	Microliter
%	Percent
±, +/-	Plus or minus
®	Registered
16S	Subunit 16
16S rRNA	Ribosomal ribonucleic acid 16S subunit
AP	Acid Phosphatase
AUC	Area under the curve
bp	Base pair
^{o}C	Degrees Celsius
c	Class
ChiSq	Chi-squared statistic
CI	Confidence interval
cm	Centimeter
CODIS	Combined DNA Index System

C_T	Cycle threshold
df	Degrees of freedom
DFSC	Defense Forensic Science Center
DNA	Deoxyribonucleic acid
et al.	And others
f	Family
F	F-statistic
FBI	Federal Bureau of Investigation
FMT	Fecal microbiome transplant
FR	Frozen
FTA	Flinders Technology Associates
GA	Georgia
HCl	Hydrochloric acid
HMP	Human Microbiome Project
i.e.	That is
IRB	Institutional review board
k	Kingdom
log2	logarithm base 2
ml	Milliliter
mtDNA	Mitochondrial deoxyribonucleic acid
n	Sample size
N	Normal
N/A	Not applicable

NaOH	Sodium hydroxide
ng	Nanogram
NGS	Next generation sequencing
NIH	National Institutes of Health
nM	Nanomolar
NPV	Negative predictive value
o	Order
OOB	Out-of-bag
OoJP	Office of Justice Programs
OTU	Operational taxonomic unit
p	p-value
p	Phylum
P30	Prostate antigen P30
PC	Principal components
PC1	Principal components axis 1
PC2	Principal components axis 2
PC3	Principal components axis 3
PCoA	Principal co-ordinate analysis
PCR	Polymerase chain reaction
pM	Picomolar
Q	Quality scores
Q1	QIIME 1
QAS	Quality Assurance Standard

qPCR	Quantitative polymerase chain reaction
R	The R Project for Statistical Computing
R^2, R^2	Effect size
RF	Refrigerated
ROC	Receiver operating characteristic
rRNA	Ribosomal ribonucleic acid
RT	Room temperature
RUO	Research use only
STD	Sexually transmitted disease
Std. Dev.	Standard deviation
STR	Short tandem repeat
SWGDAM	Scientific Working Group on DNA Analysis Methods
U.S.	United States
V3	Variable region 3
V4	Variable region 4
vs	Versus
WGS	Whole genome sequencing
Y-STR	Y-chromosome short tandem repeat

SUMMARY

The microbiome, specifically the human microbiome, is an area of intense research with potential forensic applications. If the microbiome of a given body area is individualizing, then the transfer of microbiomes between individuals could reveal contact between a victim and suspect. An important application would be in sexual assault cases where no semen is detected or analyzable. This thesis aimed to address whether the identification of individuals through their pubic hair/pubic area microbiome would support detection of biological transfer.

Sequencing of 16S rRNA gene fragments derived from pubic hair and pubic mound swabs of 43 individuals, including 12 sexually active couples and 19 unpaired singles, each providing 2 to 5 sets of samples over two months, allowed me to evaluate under which circumstances forensic individualization may be feasible. Random Forest modeling demonstrated 90% accuracy in assigning replicate samples to individuals, but had variable accuracy in pairing sexually active couples. Hierarchical clustering also provided some support for couples being more similar than random pairs of individuals, and a trend toward greater sharing with higher sexual activity was detected. Couples who never reported sexual activity in the week preceding sample collection did not cluster together, whereas couples who reported at least one instance of sexual activity in the week preceding sample collection shared clusters at least 25% of the time.

Unexpectedly, elbow and ear swabs were as similar within couples as pubic samples, although they clustered irrespective of the degree of sexual activity.

These results suggest that detection of microbiome transfer during sexual activity is not guaranteed and will not generally identify the source from the general population. To explore the potential to identify or exclude suspects from a small pool of candidates, I performed *in silico* mixture modeling of pseudo-couples. These analyses implied that a minimum of 5% contribution from a donor is required to be able to detect any change in the host's microbiome and at least 25% contribution is required for reliable detection. Evaluation of the SourceTracker tool for prediction of contributors to a microbiome profile consisting of Deblur OTU assignments also showed promise both for correctly identifying two or more contributors to a sample and for exclusion of non-contributors. While further work will be needed to refine these models, their potential is demonstrated for questions such as identification of a suspect from the general population, identification of a culprit from a small pool of suspects, and exculpation of a falsely accused individual from a small pool of suspects.

Additional limitations and concerns, such as best practices for analysis of low biomass samples, how most appropriately to handle contamination, and other crime laboratory considerations, are addressed in the concluding chapter. This thesis shows the potential of the pubic hair/pubic area microbiome and represents the first steps towards the development of a new tool for crime laboratories.

CHAPTER 1

INTRODUCTION

1.1 Investigation of Sexual Assaults

Sexual assaults are a pervasive issue in today's society with the investigation and prosecution of sexual assaults playing an important role. According to the World Health Organization, over a third of women experience partner violence or non-partner sexual violence over their lifetimes [1]. Over 284,000 Americans 12 years of age and older were sexually assaulted or raped in 2014 [2]. In the U.S. Military, there were 6083 victims in reports of sexual assault in fiscal year 2015 [3]. Within the Defense Forensic Science Center, sexual assaults make up 34% of the forensic case load (M. Hill, personal communication). Approximately 30% of rapes are reported with just 1% referred to prosecutors and <1% resulting in conviction [4, 5]. While many factors play a role in the successful prosecution of a sexual assault case, collection and examination of evidence in cases of sexual assault are very important steps in the resolution of these cases.

While evidence submitted for biological analysis can never answer the question of whether the sexual contact was consensual, the evidence can be screened for bodily fluids and DNA to support the claims that the contact occurred. The ability of forensic testing to detect sexual activity with conventional DNA typing methods

is excellent when semen has been left behind on intimate items of evidence. Further testing is limited in the absence of semen, such as when a condom has been used, there has been withdrawal before ejaculation, and when the suspect is aspermic. In some cases, Y-chromosome short tandem repeat (Y-STR) testing may be utilized where there is a low level of male DNA or the high presence of female DNA in a mixture. However, Y-STR typing is limited as it is not individualizing due to its patrilineal inheritance. There may be other biological material available, such as saliva or skin transfer, but is often limited in its ability to be linked to sexual activity. Development of techniques with the ability to utilize all available evidence assists with a thorough investigation.

Pubic hair is an evidence type with the potential to associate individuals who have had intimate contact. Currently, pubic hair evidence is not commonly tested in crime laboratories due to limited resources. Microscopic comparison of pubic hairs requires specialized training. Additionally, the association of individuals by using hair characteristics lacks vigorous scientific validation [6]. In order to perform microscopic hair comparisons, a representative sample of hairs from an individual is required. This sample includes hairs with roots, necessitating pulling of hairs from a victim already traumatized by an assault. A root must also be present on a hair in order to conduct traditional nuclear DNA typing. Human pubic hair/pubic area microbiome has the potential to identify and individual and detect mixtures of individuals when semen is not present as well as utilizing cut hairs, reducing further trauma to the victim. The pubic hair/pubic area microbiome may also be

used to eliminate a suspect, both in situations where there is a single suspect as well as a pool of suspects.

The focus of the research here was to begin the development of a new technique using the microbiome for the analysis of pubic hair/pubic area swabs for the identifying individuals and for the detection of sexual activity.

1.2 The Microbiome

The microbiome is a fascinating environment largely underappreciated by the population at-large but with impacts large and small across all types of habitats. This work aims to explore a new forensic use of the human pubic hair/pubic area microbiome by showing the individuality of this microbiome and its potential to associate couples and mixtures of microbiomes after sexual contact. A microbiome consists of the microscopic organisms, bacteria, archaea, viruses and eukaryotes, that make up a habitat. To fully understand the microbiome, one must study how the species that make up the microbiome interact with each other as well as how they interact with their host environment. Due to the difficulty in detecting and identifying the vast majority of microorganisms, much of the microbiome had been a mystery until relatively recently. Until advances in molecular biology techniques such as massively parallel sequencing, much of our knowledge regarding microbiomes had been limited to those species that could be

cloned in the laboratory. Now, massive efforts are underway to catalog the microbiome on scales large and small, from the Earth Microbiome Project [7] to the built environment of the International Space Station [8]. As it relates to health and disease, study of the human microbiome has been a rapidly growing field of study. While early studies, like the beginnings of the Human Microbiome Project [9], concentrated on cataloging the members of the microbiome, there is now a higher emphasis on the function of the microbiome and the interactions between its members and between the microbiome and its host.

There is little doubt that a healthy microbiome is essential for human health. It has been observed as far back as 1915 that germ-free animals are not as healthy as their non-sterile counterparts [10]. But what constitutes a healthy microbiome and if a dysbiotic microbiome is the cause or effect of disease remains to be resolved as well as what therapeutic interventions may be most beneficial. Changes in the microbiome have been implicated in autoimmune disorders, psychiatric disease, and others with microbiome transplants suggested as a potential therapy for some [11]. An example of successful microbiome therapy is the treatment of *Clolostrum difficile* infections using fecal microbiome transplants [12].

Multiple studies have shown that an altered microbiome is found in obese humans and mice. Alterations in diet and transfer of the fecal microbiome from obese mice to lean mice has been shown to be sufficient to induce obesity in the lean mice [13]. Studies like these illustrate a direct effect of an altered microbiome.

Identifying what makes up a “healthy” microbiome and how to return it to that state, whether by transfer of entire microbiomes from healthy individuals or inoculation with key members of the microbiome community, continues to be studied.

Multiple studies have shown that individuals sharing households have more similar microbiomes than unrelated individuals [14, 15]. This similarity even extends to the household pets. Is this similarity from contact or from sharing a built environment? When sampling multiple skin sites, sharing between couples is not consistent at all skin sites. Skin sites that show evidence of sharing include the hands and the soles of the feet. Given that some of these sites are expected not to experience direct contact, it may be that the shared environment is playing a role in the exchange of microbiome. It has been shown that a household’s microbial community rapidly becomes similar to the previous household microbial community when that family moves [16]. There are indications that the individuals may become more like each other from direct contact also. Incidental results from Tridico et al. suggested a detectable transfer of pubic hair microbiome between individuals who had intercourse [17]. Roller derby teams showed a shift towards more similar skin microbiomes after competing [18]. Further refinement of studies are needed to address the detection of direct transfer as opposed to general sharing from co-habitation.

A strength of the microbiome is its potential to differentiate individuals [19]. The human microbiome has distinct membership depending on the body area sampled with further separation of individuals within a body area [20]. Depending on the body area, the microbiome has been found to be stable over time with varying ability to return to its original state after perturbation. The gut microbiome has been shown to be remarkably resilient [21]. In the case of skin microbiomes, the difference between sebaceous versus dry skin sites plays a role in the stability of the microbiome post perturbation [22]. This stability is useful for associating samples from the same individual and for detecting when perturbations have occurred.

These characteristics of individualization, stability, and potential for detection of transfer events are desirable in a potential analysis method with forensic implications. Studies have been performed exploring the ability to link individuals and objects through their shared microbiomes [23-26]. The microbiome of an environment has been used to predict characteristics of the inhabitants of that environment, such as number of inhabitants, gender of inhabitants, and whether or not a pet lived in the environment [14-16, 27]. Beyond using the microbiome for linking individuals, environments, and/or objects, investigations are also being conducted into using the microbiome as a post-mortem interval clock [28-32]. All of these studies are preliminary and the techniques have yet to be brought into use in criminal cases.

1.3 Techniques

There are multiple analysis techniques available for investigating the microbiome. Decisions must be made at each step of the process and which technique is chosen is driven by what questions about the microbiome are being asked. Are we only interested in which species are present in the microbiome or do we also want to know what the functions of the various species are in the community? Function may be inferred by what genes are present, what genes are being transcribed, and/or what proteins are present in the community. Is genus-level identification sufficient or do we need to know species or sub-species information? The microbiome consists of any microscopic organism: bacteria, archaea, viruses, microscopic eukaryotes. Often, studies concentrate on just one category of microorganism.

Sequencing of the 16S ribosomal RNA (rRNA) gene has largely been used for basic population surveys of bacteria in a given microbiome. The 16S rRNA gene is an essential housekeeping gene present in all bacteria. We are able to determine which bacteria are present in a sample to the genus level by comparing the sequencing data from the 16S rRNA gene to a database, such as Greengenes [33, 34]. Sequences that are different by more than 97% are considered to identify different species. However, sequences with 97% or greater sequence similarity may be from different species, limiting identification mainly to the genus level.

Within the 16s rRNA gene there are eight variable regions. These regions are used for targeted amplification and sequencing using massively parallel sequencing. The choice of variable region and primer set for amplifying that region affects how successful you are in detecting all of the bacteria in the sample [35]. Some primer sets preferentially amplify certain species over others. A chosen primer set should be able to amplify as many species as possible. Sequencing of the 16S rRNA gene is well-established with databases available for genus/species annotation. The functional capability of the microbiome can be inferred from what is known about member bacteria but this data does not tell us which bacteria are active when and which bacteria present actually have particular genes. This information is also dependent on what is already known about the various species. Fine-level species differentiation typically is not possible with this method. Resolution of sub-OTU level differences may be achieved with a tool such as Deblur [36], which uses statistical methods to obtain single nucleotide resolution of amplicons. This resolution is important for the comparison of closely-related communities.

In order to determine the population of a microbiome at the species or sub-species level one must consider whole-genome sequencing (WGS). With WGS, variation present in the whole genome may be considered for species determinations. Here genes are directly sequenced so that their presence does

not need to be inferred from genus/species identifications. Gene identifications may be limited or inferred from homologous genes in other species.

Transcriptomics and proteomics are required to determine the functional activity of a microbiome. Transcriptomics will provide information on which genes are actively being transcribed while proteomics will provide information on which proteins are being produced. These tests allow for a snapshot of the community activity. Unless coupled with genus/species annotations and gene content knowledge for those genus/species, one would not know which bacteria are active and producing the transcripts or proteins. While thinking of the microbiome as a single entity may be useful for understanding its overall function in an ecosystem, it is still important to understand the function and role of individual microbiome members and how changes in those members affects both intra-microbiome and microbiome-host interactions.

1.4 The Microbiome and Forensic Science

In its most basic form, forensic science is the application of science to matters of the law. Many of the questions crime labs are asked to answer when analyzing evidence revolve around associating an individual with a crime through contact with another individual or a scene. These questions may be answered by latent fingerprint examinations, trace evidence examinations, questioned document

examinations, firearms examinations, or serology and DNA examinations, to name a few. Many of the tests performed in the crime lab have remained fundamentally the same for many years. Whatever the technique, the same basic need to make connections using the evidence remains. When current methodologies fail to provide answers to the questions at hand, new techniques may provide new avenues to answering these questions.

Analysis of the microbiome has shown promise in answering the question of linking an individual to another object or individual. Studies largely involve the development of the microbiome profile using 16S rRNA gene sequencing, as the composition of the microbiome in these instances is of more importance than the functional capability of the microbiome. Individuals have been linked back to their keyboards, mice, cell phones, and shoes [25, 37]. We are primarily interested in the skin microbiome in these instances.

As a potential new forensic technique, many questions need to be asked before the microbiome can be used in forensic analyses. These questions include the stability of the microbiome, both over time in an individual and after it has been transferred to another location, the population variation of a given microbiome, the similarity of a given microbiome for individuals in a shared environment, and how to detect the transfer of a given microbiome. Once foundational knowledge about the microbiome has been established, specific methodologies can be developed for translation to the crime laboratory. Introduction of a new DNA methodology to

an accredited crime lab is governed by the FBI's Quality Assurance Standards for Casework Laboratories [38]. While these standards specifically cover human DNA analysis, the framework of method validation may be useful in shaping the types of studies needed in considering the microbiome.

Of particular interest in this work is the pubic hair microbiome. Preliminary observations by Tridico et al. indicated transfer of the pubic hair microbiome after sexual intercourse [17]. While the sample size was small and this finding was incidental to the main topic of the paper, the ability to detect the transfer of pubic hair microbiome after sexual intercourse would be useful in cases of sexual assault where there was no transfer of semen. In this work, the aim is to lay the groundwork for using the pubic hair/pubic area microbiome as a forensic tool in establishing the identity of a person and detecting transfer of the microbiome during sexual intercourse. First, how the pubic hair microbiome is affected by storage time and storage temperature under typical crime laboratory conditions was explored. Then, a population survey of the pubic hair/pubic area microbiome was conducted. This survey was accompanied by behavioral questions in order to begin understanding how certain behaviors affect the pubic area/pubic area microbiome. Samples were collected from individuals over time also to determine the stability of the microbiome. Additionally, couples were asked to participate so that preliminary data on how the microbiome is affected by sexual activity could be assessed.

As is true in many forensic samples, there is a challenge with interpretation of mixtures. With the mixture interpretation, it is important to define the question you are attempting to answer, particularly when using a Bayesian framework for interpretation. The sequences collected during the population survey were used to make new *in silico* mixtures of known ratios to explore the limits of mixture interpretation. Lessons learned from these simulations were applied to the original sequences to show how well these models worked with actual couples. Work with the simulated mixtures also help to inform the direction of further studies that would help refine and tune possible mixture interpretation models.

A final discussion addresses the challenges with the analysis of microbiome and some of the work that remains to be done before the pubic hair/pubic area microbiome can be used in a crime laboratory. These challenges are not insurmountable and answering them will lead to a promising new forensic method utilizing new techniques with an underutilized resource, the pubic hair.

1.5 References

[1] World Health Organization. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. 2013.

[2] Truman JL, Langton L. Criminal Victimization, 2014. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. 2015.

[3] Department of Defense. Department of Defense Annual Report on Sexual Assault in the Military. Fiscal Year 2015.

[4] National Crime Victimization Survey, 2010-2014. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. 2015.

[5] National Incident-Based Reporting System, 2012-2014. Federal Bureau of Investigation. 2015.

[6] President's Council of Advisors on Science and Technology. Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. 2016.

[7] Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. BMC Biology. 2014;12:69.

[8] Venkateswaran K, Vaishampayan P, Cisneros J, Pierson DL, Rogers SO, Perry J. International Space Station environmental microbiome - microbial inventories of ISS filter debris. Appl Microbiol Biotechnol. 2014;98:6453-66.

[9] The NIH HMP Working Group. The NIH Human Microbiome Project. Genome Research. 2009;19:2317-23.

[10] Kendall A. The bacteria of the intestinal tract of man. Science. 1915;42:209.

[11] Mimee M, Citorik RJ, Lu TK. Microbiome therapeutics - Advances and challenges. Advanced Drug Delivery Reviews. 2016;105:44-54.

[12] Smits LP, Bouter KE, De Vos WM, Borody TJ, Nieuwdorp M. Therapeutic Potential of Fecal Microbiota Transplantation. Gastroenterology. 2013;145:946-53.

[13] Turnbaugh PJ, Ridaura VK, J FJ, Rey FE, Knight R, Gordon JI. The Effect of Diet on the Human Gut Microbiome: A Metagenomic Analysis in Humanized Gnotobiotic Mice. Science Translational Medicine. 2009;1:6ra14.

[14] Song SJ, Lauber CL, Costello EK, Lozupone CA, Humphrey GC, Berg-Lyons D, et al. Cohabitating family members share microbiota with one another and with their dogs. eLife. 2013;2:e00458.

[15] Ross AA, Doxey AC, Neufeld JD. The skin microbiome of cohabiting couples. mSystems. 2017;2:e00043-17.

[16] Lax S, Smith DP, Hampton-Marcell J, Owens SM, Handley KM, Scott NM, et al. Longitudinal analysis of microbial interaction between humans and the indoor environment. Science. 2014;345:1048-52.

[17] Tridico S, Murray DC, Addison J, Kirkbride KP, Bunce M. Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investigative Genetics. 2014;5.

[18] Meadow JF, Bateman AC, Herkert KM, O'Connor TK, Green JL. Significant changes in the skin microbiome mediated by the sport of roller derby. PeerJ. 2013;1:e53.

[19] Califf K, Gonzalez A, Knight R, Caporaso JG. The Human Microbiome: Getting Personal. Microbe. 2014;9:410-5.

[20] Oh J, Byrd AL, Deming C, Conlan S, Program NCS, Kong HH, et al. Biogeography and individuality shape function in the human skin metagenome. Nature. 2014;514:59-64.

[21] Lozupone CA, Stombaugh JI, Gordon JI, Jansson JK, Knight R. Diversity, stability and resilience of the human gut microbiota. Nature. 2012;489.

[22] Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. Science. 2009;326:1694-7.

[23] Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. PNAS. 2010;107:6477-81.

[24] Goga H. Comparison of bacterial DNA profiles of footwear insoles and soles of feet for the forensic discrimination of footwear owners. Int J Legal Med. 2012;126:815-23.

[25] Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, et al. Forensic analysis of the microbiome of phones and shoes. Microbiome. 2015;3:1-8.

[26] Yao X, Liu W, Han J, Pei G, Tong Y, Luo Y. Analysis of microbiome DNA on frequently touched items and from palm prints. J Forensic Sci Med. 2016;2:74-7.

[27] Meadow JF, Altrichter AE, Bateman AC, Stenson J, Brown G, Green JL, et al. Humans differ in their personal microbial cloud. PeerJ. 2015;3:e1258.

[28] Carter DO, Metcalf JL, Bibat A, Knight R. Seasonal variation of postmortem microbial communities. Forensic Sci Med Pathol. 2015;11:202-7.

[29] Damaan FE, Williams DE, Layton AC. Potential use of bacterial community succession in decaying human bone for estimating postmortem interval. Journal of Forensic Sciences. 2015;60:844-50.

[30] Finley SJ, Benbow ME, Javan GT. Microbial communities associated with human decomposition and their potential use as postmortem clocks. Int J Legal Med. 2015;129:623-32.

[31] Hauther KA, Cobaugh KL, Jantz LM, Sparer TE, DeBruyn JM. Estimating time since death from postmortem human gut microbial communities. Journal of Forensic Sciences. 2015;60:1234-40.

[32] Metcalf JL, Parfrey LW, Gonzalez A, Lauber CL, Knights D, Ackermann G, et al. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. eLife. 2013;2:e01104.

[33] DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol. 2006;72:5069-72.

[34] McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutioary analyses of bacteria and archaea. The ISME Journal. 2012;6:610-8.

[35] Klindworth A, Pruesse E, Schweer T, Peplles J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research. 2013;41:1-11.

[36] Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. eSystems. 2017;2:e00191-16.

[37] Park J, Kim SJ, Lee J-A, Kim JW, Kim SB. Microbial forensic analysis of human-associated bacteria inhabiting hand surface. Forensic Science International: Genetics. 2017.

[38] Federal Bureau of Investigation. Quality Assurance Standards for Forensic DNA Testing Laboratories. effective date September 1, 2011.

CHAPTER 2

INDIVIDUALIZATION OF PUBIC HAIR BACTERIAL COMMUNITIES AND THE EFFECTS OF STORAGE TIME AND TEMPERATURE

Published in Forensic Science International: Genetics (2017) 26: 1-96.

2.1 Abstract

A potential application of microbial genetics in forensic science is detection of transfer of the pubic hair microbiome between individuals during sexual intercourse using high-throughput sequencing. In addition to the primary need to show whether the pubic hair microbiome is individualizing, one aspect that must be addressed before using the microbiome in criminal casework involves the impact of storage on the microbiome of samples recovered for forensic testing. To test the effects of short-term storage, pubic hair samples were collected from volunteers and stored at room temperature (~20 ^{o}C), refrigerated (4 ^{o}C), and frozen (-20 ^{o}C) for 1 week, 2 weeks, 4 weeks, and 6 weeks along with a baseline sample. Individual microbial profiles (R^2 = 0.69) and gender (R^2 = 0.17) were the greatest sources of variation between samples. Because of this variation, individual and gender could be predicted using Random Forests supervised classification in this sample set with an overall error rate of 2.7% $\pm$ 5.8% and 1.7% $\pm$ 5.2%, respectively. There was no statistically significant difference attributable

to time of sampling or temperature of storage within individuals. Further work on larger sample sets will quantify the temporal consistency of individual profiles and define whether it is plausible to detect transfer between sexual partners. For short-term storage ($\leq$ 6 weeks), recovery of the microbiome was not affected significantly by either storage time or temperature, suggesting that investigators and crime laboratories can use existing evidence storage methods.

2.2 Introduction

Research performed by the Human Microbiome Project (HMP) and others on the human microbiome of the human gut, oral, other skin areas, and vagina show that different body habitats have different microbiome compositions and that within a given body habitat an individuals' microbiome is more like themselves over time than the microbiome of other people [1, 2]. From these studies, it appears that the human microbiome is almost as individualizing as a person's own human genome [3]. This potential individualization could be applied to forensic investigations.

Forensic applications of the microbiome have thus far been primarily limited to touch-type samples [4, 5], soil comparison [6-8], and determining post-mortem interval [9, 10]. The use of the human microbiome in sexual assault investigations has not been explored. In the best-case scenario, evidence of sexual contact can

be found in the form of human DNA from sperm on an intimate sample, such as a vaginal swab. However, sperm may not be found due to no sexual contact, condom use, ejaculation outside of the body, or lack of sperm production. In the absence of sperm, there is currently little that can be done to show that sexual contact has occurred. Y-chromosome short tandem repeat (Y-STR) testing alone may show the presence of male DNA, especially in the presence of high amounts of female DNA, but cannot provide information on the body fluid or body part from which it originated. Y-STR testing may also be used to detect male DNA post-coital (5-6 days) where traditional nuclear DNA typing may not [11]. In the absence of corresponding serological testing that may provide information about the presence of seminal fluid, such as acid phosphatase (AP) or P30, no conclusions may be drawn from Y-STR results regarding the body origin of the sample. Mitochondrial DNA (mtDNA) is powerful for lineage studies and is useful for single hairs which are unsuitable for nuclear DNA testing due to the low amount of nuclear DNA present in hairs [12, 13]. These techniques have great powers of exclusion, as in the case of an individual claiming to be the Boston Strangler [14]. Both Y-STR and mtDNA testing are not individualizing as they are inherited paternally and maternally, respectively, and suffer from interpretation issues when there is a mixture of DNA.

Motivating our study, Tridico et al. [15] analyzed a limited sample set (scalp and pubic hairs from 7 individuals at three time points) and found that scalp hairs

could be distinguished from pubic hairs by their microbiome, which was individualizing. The pubic hair microbiome was stable over time and indicative of gender. For two individuals in their study, their pubic hair microbiome was more similar to each other at one time point than at the other time points, which grouped together by individual. Further inquiries revealed that this cohabitating couple had sexual intercourse approximately 18 hours prior to the collection, suggesting that their microbiome had cross-transferred during this event. Additionally, the fact that this couple cohabitated did not automatically cause their microbiome to be more alike outside of sexual intercourse. Data analysis was limited to construction of principal co-ordinate analysis (PCoA) plots and core vs transient operational taxonomic unit (OTU) comparisons. The statistical significance of any differences between samples and/or individuals was not addressed. While the scalp and head hairs were distinguishable from each other microbially, the recovered biome was not compared to other body areas to determine their uniqueness within the body. Due to the limited number of participants in this study, more work needs to be done with more individuals to determine the uniqueness of an individual's pubic hair microbiome signature across multiple time points.

Additionally, other concerns such as stability over time on the body, stability over time in storage, degree of transfer between individuals, and behavioral effects on the microbiome need to be addressed before this type of information can be used

in a forensic context. Ultimately, rather than typing individual hairs from a crime scene for potential donor/assailant identification, microbiome profiling is envisaged as a novel tool for establishing transfer from samples taken from the victim and suspect. Either cut pubic hairs or swabs from the pubic mound, taken from the victim, could be profiled for presence of microbial species that are representative of another individual, potentially the assailant, in situations where nuclear DNA from the assailant cannot be obtained from the victim. Additionally, the combination of human genetic and microbiome profiles may be complementary.

Samples collected as part of a criminal investigation are stored under various conditions that may not be considered ideal. In studies on soil [16, 17], human fecal and skin [16, 18, 19], human vaginal [20], and animal fecal samples [21, 22] where samples were stored from temperatures ranging from room temperature (20 ^{o}C) to ultra-low frozen (-80 ^{o}C) within and out of storage media, results vary as to the impact of various storage conditions on the recovered microbiome. Because past storage studies had such varying results and none of the studies addressed human pubic hairs, the goal of this study was to investigate the effects that storage temperatures common to forensic laboratories (room temperature, refrigerated, and frozen) and short-term storage time have on the microbiome recovered from cut human pubic hairs. Additionally, the inter-individual and intra-

individual variance in this sample set was investigated to gain preliminary information on the utility of the pubic hair microbiome for individualization.

2.3 Methods

2.3.1 Sample Collection

Six adults (three male (SM01-SM03) and three female (SF01-SF03)) who self-identified as healthy were anonymously recruited from the Defense Forensic Science Center. As no human genetic data was to be generated and no health information could be directly derived from the microbiome data, the project was classified by the Defense Forensic Science Center Human Protections Administrator as not human subjects' research. Collection packets coded with a unique alphanumeric code (SM01-SM03, SF01-SF03) for each packet were provided for anonymous pickup and return. The coded collection packets contained a sterile suture removal kit containing scissors and forceps, sterile plastic tongs, sterile 50 mL urine collection vials, and collection instructions. Gloves were also provided so that participants could choose their appropriate size. Participants were instructed to conduct the hair collection within the same 24-hour period and to return the completed collection packet by the end of that time window. Each participant collected at least 80 cut pubic hairs prior to bathing or showering. The participants were instructed to collect from all around the

genital area with the females avoiding hairs from the labial area or hairs in direct contact with the vaginal opening. The participants were instructed to wear gloves during the collection and to use the forceps or tongs to grip and handle the hairs. The hairs were placed into the urine collection vials. The participants labeled the completed collection packet with the date and time of collection. The completed collection packets were held at room temperature until participants returned the completed collection packets to the provided drop box.

After removing hairs to perform the baseline extraction, the remaining hairs per individual were divided roughly evenly between 12 sterile 50 mL urine collection vials for storage. The hairs were stored at room temperature (RT, ~20 ^{o}C), refrigerated (RF, 4 ^{o}C), or frozen (FR, -20 ^{o}C) for 1 week, 2 weeks, 4 weeks, and 6 weeks. At least 7 cm of total combined hair length (7.1-12.6 cm, mean 8.9 cm) was used for each extraction. Hairs were cut to be approximately 1 cm in length before being put into the extraction tube.

2.3.2 DNA Extraction and Sequencing

Genomic DNA was extracted using the MO BIO PowerFecal® DNA Isolation Kit following manufacturer's directions with the modification of performing the final elution with 50 μl Solution C6. Extractions were performed at each storage time point with a reagent blank included with each extraction set. Extracts were concentrated to dryness using a vacuum concentrator and stored at -20 ^{o}C until library preparation. Immediately prior to amplification, samples were reconstituted

in 6 μl amplification-grade water and quantified using a Qubit® 2.0 Fluorometer (ThermoFisher) and Qubit® dsDNA HS Assay (ThermoFisher) with 2 μl of sample. All samples were below the detection limit of the Qubit® assay (0.50 ng/ml) or between the limit of detection and 5 ng/μl.

Sequencing libraries were prepared following the Illumina® 16S Metagenomic Sequencing Library Preparation guide [23]. Because all samples were below the recommended input of 5 ng/μl DNA, amplification was performed with 2.5 μl of each reconstituted sample on the V3 and V4 region of the 16S rRNA genes using the primers 341F and 805R from Klindworth et al. [24] with the addition of the Illumina® overhang adapter sequences. After Index PCR and clean-up, libraries were quantified with the KAPA Library Quant for Illumina® (ROX Low) (KAPA Biosystems) on an Applied Biosystems® 7500 Real-Time PCR System (ThermoFisher). Libraries were normalized to 50 pM, if possible, and 10 μl of each library was combined to form the pooled library. Four libraries (SF01 6 weeks room temperature 40.7 pM, SF03 6 weeks room temperature 27.8 pM, SM02 6 weeks refrigerated 39.7 pM, and SM03 6 weeks frozen) were below 50 pM and 10 μl of each of these libraries was added to the pooled library.

Due to the low level of recovered libraries, library denaturation was modified from the Illumina workflow, which recommends a starting library input of 4nM, with the 50 pM library pool by adjusting the concentration of sodium hydroxide (NaOH) added to the library and using hydrochloric acid (HCl) to neutralize the NaOH. The

50 pM pooled library was denatured with 1N NaOH to a final NaOH concentration of 0.1N (50 μl 1NaOH, 450 μl 50 pM library pool). After incubation at room temperature for 5 minutes, a volume of 1 N HCl equal to the amount of 1N NaOH used for denaturation (here, 50 μl) was added to neutralize the library. The pooled denatured library (500 μl) was diluted with 500 μl of pre-chilled HT1 Hybridization Buffer (Illumina) for a concentration of ~20 pM. The rest of the Illumina workflow was followed, diluting the library to a final loading concentration of 4 pM. Sequencing was carried out on a MiSeq® FGx (Illumina) in Research Use Only (RUO) mode using the MiSeq® v3 Reagent Kit (Illumina).

2.3.3 Sequence Analysis and Community Comparisons

Prior to community analyses, a series of quality filtering steps were performed. Sequences with a base quality score of less than 25 and sequences with a sequence length of less than 100 bp were removed with Trim Galore! [25] and forward and reverse reads were merged with PANDAseq Assembler [26]. Remaining analyses were conducted using QIIME [27]. Chimera checking was conducted with Usearch61 (*identify_chimeric_seqs.py*) [28, 29] and chimeras were filtered from the sequence reads. Open reference OTU picking (*pick_open_reference_otus.py*) during which sequences are clustered based on 97% similarity was performed using the default settings (UCLUST [28], PyNAST alignment [30], Greengenes 13_8 [31]). Core diversity analysis (*core_diversity_analysis.py*) was performed on the resulting OTU table at a depth of 952 reads (lowest read count in a non-reagent blank sample) to generate

taxonomy plots, UniFrac distances, and PCoA plots. The reagent blanks were largely composed of Rhizobiales, which was also present in other samples and not a skin commensal microorganism. Rhizobiales was filtered from the OTU tables for further analyses. At this point, the non-reagent blank sample with the lowest read counts had 326 read counts and the reagent blank samples ranged from 16 to 711 read counts. In order to limit information loss from rarefying at too shallow a depth while eliminating the reagent blanks, core diversity analysis was repeated at a depth of 900 reads for taxonomy and beta diversity through weighted UniFrac. At this level, all of the reagent blanks and 5 samples were filtered from the remaining analyses.

To compare the similarities and differences between samples, several metrics were calculated. Alpha diversity was calculated using the Simpson Index (*alpha_diversity.py –m simpson*) to investigate the variation in alpha diversity over the storage times and temperatures. The significance of the categories individual, gender, storage time, and storage temperature was determined with the vegan::adonis nonparametric method for multivariate analysis of variance on the weighted UniFrac distances (*compare_categories.py –method adonis*) [32]. To compare intra- and inter-individual variation in the weighted UniFrac distances, tests of significance were performed using a two-sided Student's two-sample t-test. The ability to predict individual, gender, storage time, and storage temperature was calculated through supervised learning with the Random Forests method with 10-fold cross validation and 500 trees (*supervised_learning.py –e*

cv10) [33]. Differential abundance of OTUs between genders at the order level was calculated using DESeq2 (*differential_abundance.py -a DESeq2_nbinom*) [34, 35].

2.4 Results

Six volunteers (three males, three females) from the Defense Forensic Science Center (DFSC) in Forest Park, GA, provided cut pubic hairs which were stored for 1 week, 2 weeks, 4 weeks, and 6 weeks at room temperature (~20 ^{o}C), refrigerated (4 ^{o}C), and frozen (-20 ^{o}C) prior to extraction. Hairs were also extracted within 24 hours of collection. Results from the amplification and next generation sequencing of the V3/V4 region of the 16S rRNA gene were analyzed. There were 1,571,485 reads generated that passed quality and chimera checking with an average of 18,933 reads per sample (range of 16 reads to 144,375 reads) for the 83 samples and reagent blanks. Once the non-skin commensal Rhizobiales was removed and samples with reads below 900 read counts were filtered from the data there were a total of 1,566,698 reads with an average of 21,460 reads per sample (range of 900 reads to 144,375 reads) for the remaining 73 samples. The samples with fewer than 900 read counts post quality control and filtering were from two individuals (SF02 (2) and SF03 (3)) and multiple storage conditions (room temperature/baseline, frozen/two weeks (2), room temperature/six weeks, and frozen/six weeks). There was no obvious relationship

between storage time or storage temperature and which samples failed to produce reads above 900 post-quality control filtering nor was a general trend for read counts between the genders. As no information, such as bathing habits, was collected from the participants, there is no way of knowing if any common participant behavior affected the recovery of the microbiome. The remaining samples were rarefied through core diversity analysis to 900 reads per sample in order to equilibrate diversity comparisons.

At the genus level, the taxonomy of the microbial communities of each individual can be distinguished from the rest with six distinct groupings (Figure 2.1). Comparison of the rarefied genus taxonomy to non-rarefied genus taxonomy (Supplemental Figure 2.1) shows no significant loss of information or loss of the ability to group individuals with rarefaction. The Pearson correlation coefficient between the two genus-level taxonomy tables was 0.993 (p=0.001). Weighted UniFrac distances were calculated on the rarefied samples in order to perform comparisons on the samples which take into account the genera that are present, relative abundance of those genera, and the relatedness of the genera. Using Adonis on the weighted UniFrac distances to calculate contributions to the variance, Individual and Gender (69.1% and 17.3%, respectively) were the only categories that were statistically significant ($p < 0.001$) (Table 2.1). Time and temperature were assessed within individuals with no statistically significant effects.

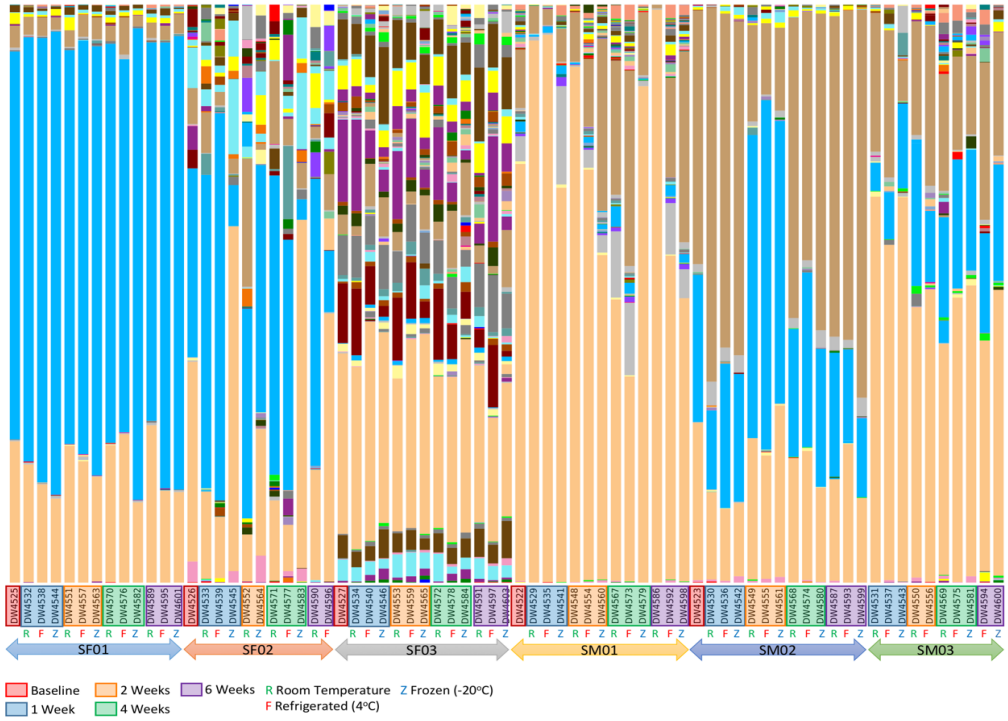


Figure 2.1 Taxonomic distribution at the genus level across storage time and temperature conditions. Sorted by individuals. Samples were normalized to 900 reads per sample.

Table 2.1 Effect sizes (R^2) calculated using the weighted UniFrac distance matrices and the vegan::adonis method. Individual and gender are grouped across all samples. Storage time and temperature are grouped within individuals.

Samples	Comparison	R^2	p-value
All	Individual	0.691	<0.001
All	Gender	0.173	<0.001
SF01	Time	0.261	0.937
SF01	Temperature	0.317	0.065
SF02	Time	0.0678	0.665
SF02	Temperature	0.263	0.198
SF03	Time	0.0864	0.324
SF03	Temperature	0.327	0.065
SM01	Time	0.112	0.244
SM01	Temperature	0.130	0.54
SM02	Time	0.0936	0.29
SM02	Temperature	0.0241	0.913
SM03	Time	0.197	0.055
SM03	Temperature	0.256	0.251

Groupings by Individual and Gender can be seen in the PCoA plots based on the weighted UniFrac distances (Figure 2.2). The same PCoA plot was colored in each panel based on individual, gender, storage time, and storage temperature. Individuals were spread along the PC1 axis, accounting for 37.5% of the variation in weighted UniFrac distances, with distinct groupings of individuals SF01, SF03,

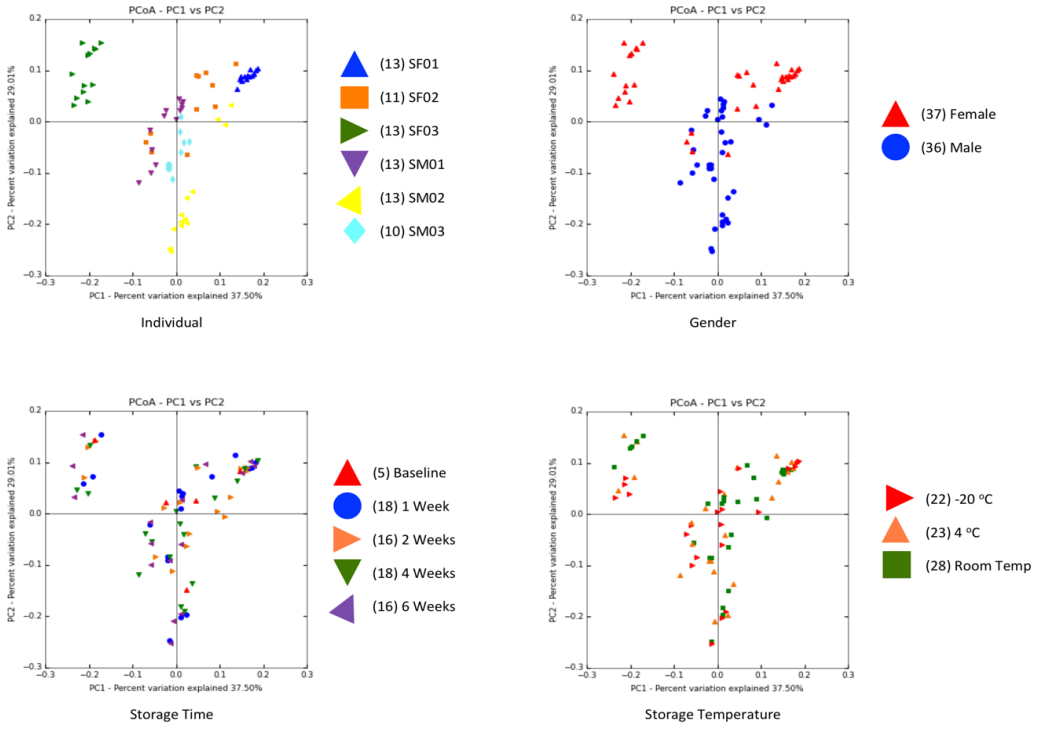


Figure 2.2 Effect of individual, gender, storage time, and storage temperature on the bacterial communities of the pubic hair. Principal component plots generated using weighted UniFrac distance matrices.

and SM02. The distinct PCoA grouping of individual SF03 is consistent with the high alpha diversity of her samples. Gender differences are spread along the PC2 axis, accounting for 29.01% of the variation in the weighted UniFrac distances. PC axes 3-10 each account for 8.73% of the variation in the sample or less (data not shown). Storage time and temperature do not have an appreciable impact on the variation in the sample.

Due to the potential for a single genus to be represented by multiple OTUs, differential abundance of OTUs between genders was analyzed at the order level. There were ten orders that were significantly differentially abundant between females and males (Table 2.2). Among these, nine of the orders were more abundant in females, with Lactobacillales and Bifidobacteriales being the most

Table 2.2 OTUs at order level which are significantly differentially abundant using DESeq2. Positive fold change represents OTUs that are more abundant in female samples. Negative fold change represents OTUs that are more abundant in male samples.

OTU	baseMean	log2FoldChange	Standard Error	Wald statistic	p-value	adjusted p-value
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales	25.5	3.88	0.391	9.94	2.72E-23	1.58E-21
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales	164	-2.36	0.245	-9.66	4.44E-22	1.29E-20
k__Bacteria;p__Actinobacteria;c__Actinobacteria;o__Bifidobacteriales	13.7	4.03	0.438	9.21	3.15E-20	6.09E-19
k__Bacteria;p__Firmicutes;c__Clostridia;o__Clostridiales	60.4	2.30	0.304	7.58	3.53E-14	5.12E-13
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales	18.4	2.78	0.401	6.94	4.00E-12	4.64E-11
k__Bacteria;p__Actinobacteria;c__Coriobacteriia;o__Coriobacteriales	2.61	1.47	0.337	4.37	1.24E-05	0.000120
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales	3.21	1.57	0.380	4.13	3.67E-05	0.000304
k__Bacteria;p__Fusobacteria;c__Fusobacteriia;o__Fusobacteriales	2.21	1.07	0.341	3.14	0.00171	0.0124
k__Bacteria;p__Proteobacteria;c__Epsilonproteobacteria;o__Campylobacterales	3.98	1.17	0.389	3.00	0.00273	0.0176
k__Bacteria;p__Cyanobacteria;c__Chloroplast;o__Streptophyta	2.24	1.05	0.359	2.92	0.00352	0.0204

differentially abundant (3.88 and 4.03 log2 fold change, respectively). Bacillales was more abundant in males than in females (2.36 log2 fold change). Within each gender, there was variable representation of each of these orders among individuals (Supplemental Table 2.1). Additionally, there was not an order that was specific to any gender.

Individuals showed markedly different levels of taxonomic variability as assessed by comparison of the weighted UniFrac distances between and within individuals as well as to the whole (Figure 2.3). The overall average distances between individuals were greater than within an individual. Individuals SF01 and SM03 had the lowest variation in weighted UniFrac distances within an individual. Individual SM02 had the largest variation in UniFrac distances within an individual. A two-sided Student's two-sample t-test was used to compare whether the weighted UniFrac distances were more alike within an individual's samples than compared to another individual's samples (Supplemental Table 2.2). Out of 15 individual versus individual comparisons, nine comparisons were statistically significant (Bonferroni-corrected p-value <0.05) for the individual's weighted UniFrac distances being more similar to himself or herself than the compared individual. All of the male to male comparisons resulted in weighted UniFrac distances that were more similar to each other than different.

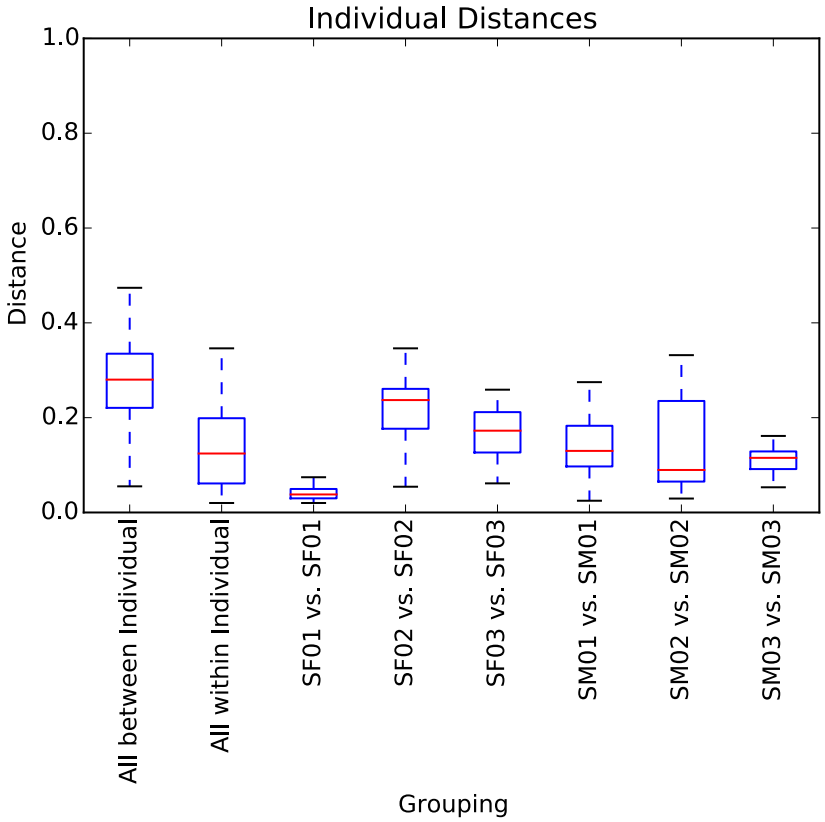


Figure 2.3 Comparison of weighted UniFrac distances within and between individuals.

The ability to predict individual identities, gender, storage time, and storage temperature was investigated using the Random Forests [33] supervised classification. Only the predictions within individual and gender performed better than random guessing (Table 2.3). Individuals were predicted with an estimated error of 0.027 $\pm$ 0.058 compared with the random guessing baseline error of 0.82. Gender was predicted with an estimated error of 0.017 $\pm$ 0.052 compared with the random guessing baseline error of 0.49. The estimated error ranges for storage

Table 2.3 Prediction of categories using Random Forests supervised classification within categories using OTUs as predictors. Values were calculated using ten-fold cross validation with 500 trees. Prediction for storage time and storage temperature predictions were performed on the data set as a whole and for each individual.

Category		Estimated error	Baseline error (for random guessing)	Ratio baseline error to observed error
Individual		0.027 +/- 0.058	0.82	31
Gender		0.017 +/- 0.053	0.49	30
Time				
	all	0.89 +/- 0.12	0.75	0.85
	SF01	1.0 +/- 0.00	0.77	0.77
	SF02	0.93 +/- 0.12	0.73	0.78
	SF03	0.75 +/- 0.25	0.77	1.0
	SM01	0.77 +/- 0.25	0.77	1.0
	SM02	0.52 +/- 0.28	0.69	1.5
	SM03	1.0 +/- 0.00	0.70	0.70
Temperature				
	all	0.72 +/- 0.12	0.62	0.86
	SF01	0.60 +/- 0.28	0.62	1.0
	SF02	0.33 +/- 0.33	0.55	1.6
	SF03	0.60 +/- 0.28	0.62	1.0
	SM01	0.67 +/- 0.41	0.62	0.92
	SM02	0.87 +/- 0.18	0.62	0.71
	SM03	1.0 +/- 0.00	0.60	0.60

time and temperature for the entire data set and within each individual were 0.52 $\pm$ 0.28 to 1.0 $\pm$ 0.00 and 0.33 $\pm$ 0.33 to 1.0 $\pm$ 0.00, respectively, which was approximately the same as the random guessing baseline error. For 71 out of 73 samples, the cross-validation label probabilities for individuals was highest for the correct individual (Supplemental Table 2.3), leading to the correct prediction of the individual. Individual SF03 had the highest probabilities for the correct individual label. This result agrees with the PCoA plot where SF03 is clearly delineated from

the remaining samples. The two samples that were assigned incorrect individual labels through supervised classification came from two different individuals, SF02 and SM02, so that those individuals were correctly identified 10/11 and 12/13 times, respectively.

To further assess the impact of storage time and temperature on the samples, the alpha diversity, a metric for the number of different types of individuals in a population, of the samples was determined with the Simpson Index where 0 is no diversity and 1 is infinite diversity. There was a wide variety of alpha diversity between individuals that was not dependent on storage time or storage temperature (Figure 2.4). The average alpha diversity across all samples was 0.91 ± 0.048 with a range of 0.84 ± 0.024 (SF01) to 0.98 ± 0.0027 (SF03). At each storage temperature there was no consistent trend across storage time of increasing or decreasing alpha diversity. Additionally, the alpha diversity of the samples was not consistently affected by the storage temperature.

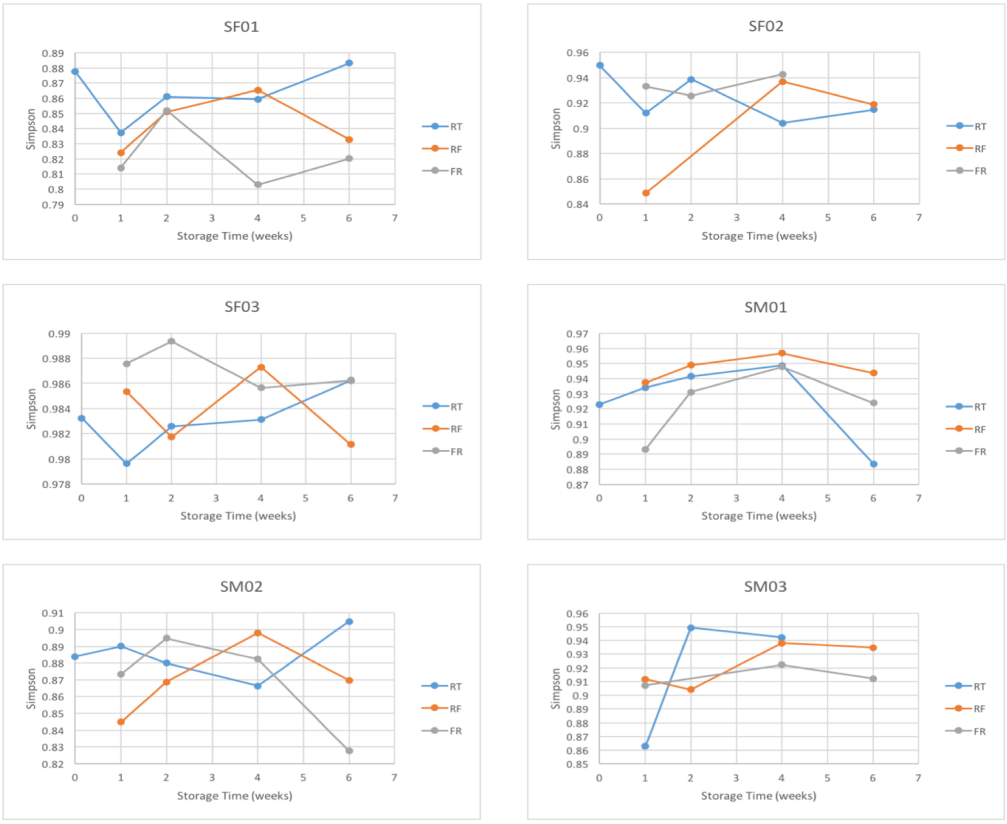


Figure 2.4 Comparison of Simpson Index of alpha diversity for each individual across storage time for each storage temperature. The higher the Simpson Index, the more diverse the sample.

2.5 Discussion

The largest sources of variation in these samples were the individuals themselves and gender. The ability to identify the individual from whom the sample originated performed well overall; however, the probability of predicting the correct individual was influenced by the variation within and between individuals. While there was no OTU at the order level that was specific to a gender, there were some orders

which were more abundant in one gender than the other. However, heterogeneity among women in the variance of gender-biased taxa limit their use as diagnostic markers for gender. Samples from more individuals of both genders will provide more information on the inter- and -intra gender specificity of these orders.

Because all of the hairs were collected from an individual at the same time, there may be some person-dependent hair-to-hair variation contributing to the differences within an individual, although Tridico et al. [15] observed longitudinal consistency of individual pubic hair samples. Additionally, the presence of a mixture of individuals in a given sample, as in from sexual contact, is unknown. One woman, SF03, had considerable greater diversity than the others, but there is little in her profile to suggest that this is because of mixing, and more controlled analysis of partners' profiles will be needed to resolve whether transfer as a result of sexual contact can be detected. The amount of hair used for each sample was small but, in most cases, the individuals were distinguishable from one another. While the read counts were generally low and rarefying the reads did not have a significant effect on the recovered taxonomy, increasing the amount of hair used in an extraction may reduce the intra-individual variability seen.

Our sampling included up to 12 samples per individual, so incorporates replicate extraction, but it would nevertheless also be preferable to perform replicate extractions of each type to quantify variation especially given the low biomass samples of used here. Multiple extractions at each sampling point was not done in

this study due to sample collection limitations in order to not burden the study participants as all hair collected from each participant was done at a single time point. Generally, this will not be possible in forensic settings either. Other studies have shown that sequencing results are mildly affected by several steps of the sample preparation process, from extraction through library preparation and sequencing [36, 37]. Despite this source of variability, technical replicates [37] and libraries of the same sample prepared with different primers [38] retain their overall similarity. If variation caused by sample preparation was significant, we would be less likely to be able to identify the individual from which a sample originated, which was not the case here. Given the low error rates for Illumina platforms (below 0.4%) [39] and use of 97% identity threshold for OTU classification, single nucleotide errors caused by the sequencing process itself should not affect the OTU classification [38]. Here, hairs were randomly divided from single collections per individual for storage and then extracted. All of the samples of each time point were extracted simultaneously, and all samples were amplified, indexed, and sequenced in a single batch, which should also reduce the impact of technical sources of variation. Optimization of the amount of hair to use and how many replicates to perform, in order to satisfactorily capture intra-individual variation for a single collection of hair, will likely emerge as various investigators and labs contract results. We expect this source of variability to be less than temporal variation, but large datasets will inform future guidelines for the collection and analysis of the pubic hair microbiome. However, as much forensic science evidence involves limited amounts of sample, perhaps as small as one

hair, one must always keep in mind that optimal amounts of sample are not always available.

Because it is rare for forensic evidence to be analyzed soon after it has been collected and storage conditions are variable from agency to agency, it was important to investigate what affect various storage times and temperatures have on the pubic hair microbiome. Thus far, studies on storage of microbiome from various other sources have been inconsistent in their results. Soil studies concluded that the origin of the soil samples had a bigger impact on microbiome composition than time or temperature of storage [16, 17]. The method of preservation (freezing, immersion in ethanol, application to FTA cards, and immersion in RNAlater) had a greater effect on spider monkey fecal samples than did storage time [21]. Overall microbial population-based differences in feline and canine microbiota stored under short term refrigeration were limited, though the authors noted potential changes in individual genera [22]. While Lauber et al. [16] found some variation in some human fecal taxa under different storage conditions, overall phylogeny was not affected by storage time or temperature in the fecal or skin samples. However, Flores et al. [18] and Gorzelak et al. [19] did report changes in bacterial diversity over storage time and temperature though they disagree on the use of RNAlater for the storage of samples. Bai et al. [20] found no significant differences in samples stored for varying storage conditions.

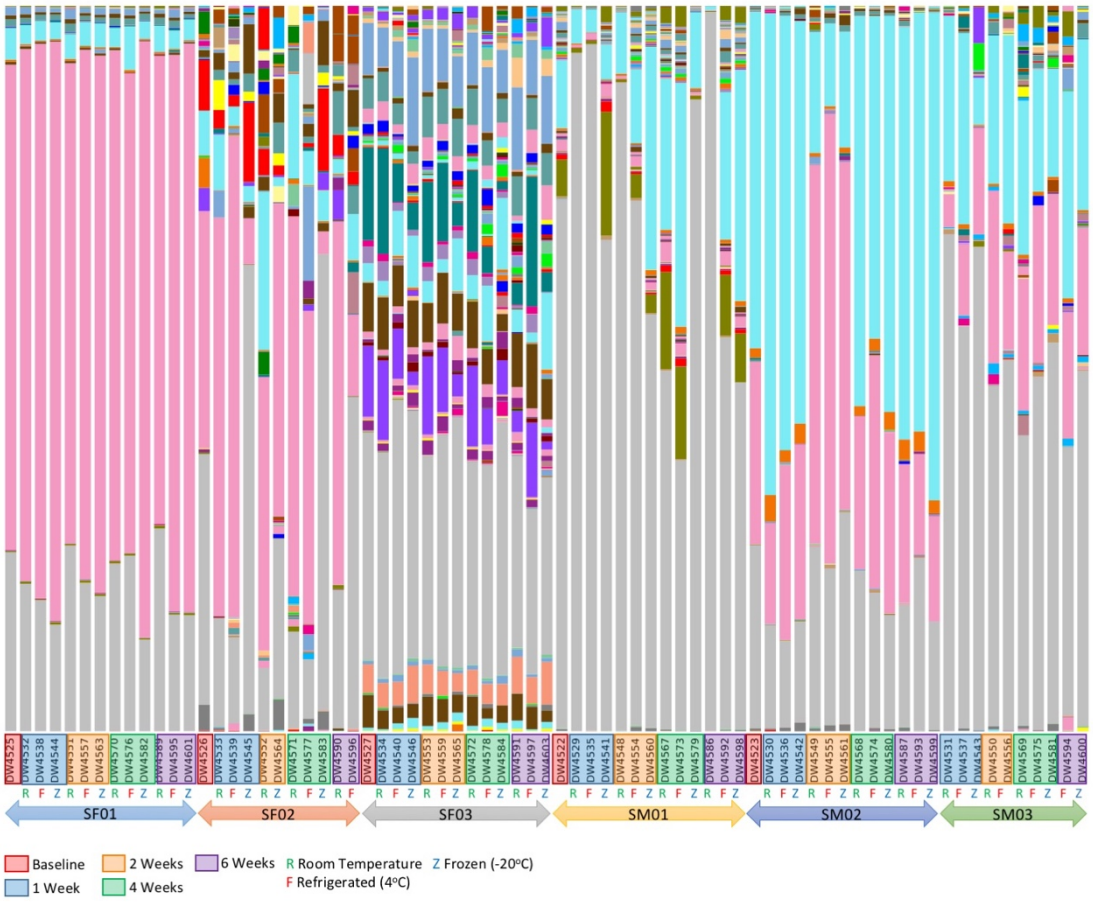
In this study, the effects of storage time and temperature on the recovered microbiome were random and did not appear to play a role in the taxonomic profiles. The alpha diversity, which accounts for the diversity within a community through the number of species in that community, was variable within an individual for each sample but did not correspond to the storage time or temperature. The visualization of the beta diversity (which accounts for the diversity between communities using the identities of the species in each community) with PCoA plots and by calculation of effect sizes also confirms that storage time and temperature did not play a significant role in causing variation in these samples. Technical noise and between-hair differences are likely to contribute to the residual intra-individual variability, but are a small component of the variation relative to inter-individual differences.

The use of the pubic hair microbiome shows promise for use in forensic investigations. However, the number of individuals tested here was small, so more work will need to be done on a larger population size across multiple time points to determine how differentiated the human pubic hair microbiome is both between individuals and between body sites. Given that pubic hair may not be present on all individuals, the microbiome of this area without hair will also need to be investigated. Once it can be robustly demonstrated that the inter-individual variation in pubic area microbiome is greater than the intra-individual variation, additional studies such as the rates of microbiome transfer between two or more individuals during intimate contact and the persistence of transferred microbiome,

as well as longer-term storage studies will also need to be conducted prior to use in criminal investigations.

While outside the scope of this study, the issue of mixtures of microbiomes from multiple individuals is one that will need to be addressed. A published study [40] of samples collected from the vaginas and penile skin of heterosexual partners where some female partners had bacterial vaginosis found that the microbiome from the male partners in the bacterial vaginosis pairs was more similar to that from their partner than from another female with bacterial vaginosis. However, this finding was not true in the partner's where the female did not have bacterial vaginosis. In these cases, the male partner's microbiome did not look more like their partner's than another female without bacterial vaginosis. Further research in progress will address questions of how to tell that a microbiome sample represents a mixed sample, and what types of reference datasets will be needed to facilitate comparisons of microbiomes. It is certainly unlikely that there will be a database of potential public hair microbiomes available for comparison, but intuitively it is plausible that once a potential perpetrator has been identified, it should be possible to ask whether the observed forensic profile could have derived from his pubic area. Statistics will need to be developed to establish the appropriate probabilities with reference to typical inter-individual variability. A related question is whether metagenomics sequencing may provide higher resolution by adding data on strain specificity to complement the inference of OTU abundance from 16S sequence alone.

2.6 Supplemental Figures and Tables



Supplemental Figure 2.1: Taxonomic distribution at the genus level without rarefaction across storage time and temperature conditions. Sorted by individuals.

Supplemental Table 2.1 Read counts for OTUs at the gender level which are differentially abundant across genders

Female

OTU	DW4525	DW4532	DW4538	DW4544	DW4551	DW4557	DW4563	DW4570	DW4576	DW4582	DW4589	DW4595	DW4601	DW4526	DW4533	DW4539	DW4545	DW4552	DW4564	DW4571	DW4577	DW4583	DW4590	DW4596	DW4527	DW4534	DW4540	DW4546	DW4553	DW4559	DW4565	DW4572	DW4578	DW4584	DW4591	DW4597	DW4603
k__Bacteria p__Actinobacteria c__Actinobacteria o__Bifidobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	3	0	1	0	0	9	0	1	1	93	104	60	17	98	88	26	111	54	42	28	97	9
k__Bacteria p__Actinobacteria c__Coriobacteriia o__Coriobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	10	7	9	8	9	4	8	2	15	4	10	10
k__Bacteria p__Bacteroidetes c__Bacteroidia o__Bacteroidales	2	0	2	1	3	3	3	4	0	3	2	3	1	0	0	0	13	0	2	0	8	15	2	0	73	72	74	91	64	81	106	67	80	55	99	99	79
k__Bacteria p__Cyanobacteria c__Chloroplast o__Streptophyta	0	0	0	0	0	0	5	0	0	0	0	0	2	0	33	6	2	0	0	3	114	0	3	0	0	0	5	0	0	0	0	0	0	8	0	0	3
k__Bacteria p__Firmicutes c__Bacilli o__Bacillales	38	30	30	21	28	37	36	24	65	17	28	23	31	57	64	24	24	201	9	136	42	42	19	153	31	15	71	64	30	19	55	29	95	85	27	13	96
k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales	2	0	0	2	0	1	0	2	0	0	0	4	1	63	36	11	124	0	16	2	10	139	30	22	149	168	52	47	131	113	40	147	45	21	37	203	25
k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales	30	15	14	15	36	14	19	25	18	17	20	23	9	12	75	43	103	19	98	31	32	102	68	105	168	158	160	210	203	160	191	175	225	174	293	179	276
k__Bacteria p__Fusobacteria c__Fusobacteriia o__Fusobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0	5	0	2	1	5	1	6	0	3	1	2	5	0	5	0	4	0
k__Bacteria p__Proteobacteria c__Epsilonproteobacteria o__Campylobacterales	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	11	0	1	0	1	0	0	2	10	6	8	13	14	26	10	5	3	7	7	33
k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Enterobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	39	0	4	3	27	17	24	0	0	0	0	0	0	1	9	0	0	0	0	0	8	0	2	0

Male

OTU	DW4522	DW4529	DW4535	DW4541	DW4548	DW4554	DW4560	DW4567	DW4573	DW4579	DW4586	DW4592	DW4598	DW4523	DW4530	DW4536	DW4542	DW4549	DW4555	DW4561	DW4568	DW4574	DW4580	DW4587	DW4593	DW4599	DW4531	DW4537	DW4543	DW4550	DW4556	DW4569	DW4575	DW4581	DW4594	DW4600
k__Bacteria p__Actinobacteria c__Actinobacteria o__Bifidobacteriales	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	3	0
k__Bacteria p__Actinobacteria c__Coriobacteriia o__Coriobacteriales	3	0	0	2	0	1	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k__Bacteria p__Bacteroidetes c__Bacteroidia o__Bacteroidales	13	1	1	0	0	12	10	10	8	2	0	11	7	3	5	0	0	0	6	2	0	6	2	1	1	2	0	0	1	0	0	0	0	2	0	0
k__Bacteria p__Cyanobacteria c__Chloroplast o__Streptophyta	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
k__Bacteria p__Firmicutes c__Bacilli o__Bacillales	96	44	26	53	78	106	278	214	314	101	6	155	307	409	619	541	547	160	114	140	485	400	520	520	526	633	222	252	65	184	288	188	158	125	253	205
k__Bacteria p__Firmicutes c__Bacilli o__Lactobacillales	3	0	3	0	1	10	2	7	2	2	0	3	3	1	1	0	1	0	0	0	1	0	0	0	0	1	3	0	2	2	0	4	1	3	1	1
k__Bacteria p__Firmicutes c__Clostridia o__Clostridiales	50	2	2	11	4	41	33	45	58	7	0	73	40	6	3	9	22	3	16	24	17	5	8	8	7	2	5	7	6	11	1	13	12	15	23	33
k__Bacteria p__Fusobacteria c__Fusobacteriia o__Fusobacteriales	1	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0
k__Bacteria p__Proteobacteria c__Epsilonproteobacteria o__Campylobacterales	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0
k__Bacteria p__Proteobacteria c__Gammaproteobacteria o__Enterobacteriales	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	3	0	0	0	0	3	5	0	2	0	0

Supplemental Table 2.2: Comparison of similarity between weighted UniFrac distances within an individual's samples and another individual's samples using a two-sided Student's two-sample t-test

Group 1	Group 2	t statistic	Parametric p-value	Parametric p-value (Bonferroni-corrected)
SF01 vs. SF01	SF02 vs. SF02	-26.0	1.64E-53	4.60E-52
SF01 vs. SF01	SF03 vs. SF03	-19.3	4.58E-43	1.28E-41
SF01 vs. SF01	SM01 vs. SM01	-14.4	2.17E-30	6.08E-29
SF01 vs. SF01	SM02 vs. SM02	-9.30	1.32E-16	3.70E-15
SF01 vs. SF01	SM03 vs. SM03	-20.8	1.10E-41	3.08E-40
SF02 vs. SF02	SF03 vs. SF03	5.96	2.15E-08	6.03E-07
SF02 vs. SF02	SM01 vs. SM01	8.09	3.54E-13	9.90E-12
SF02 vs. SF02	SM02 vs. SM02	5.91	2.86E-08	8.00E-07
SF02 vs. SF02	SM03 vs. SM03	11.7	0	0
SF03 vs. SF03	SM01 vs. SM01	2.72	0.00727	0.203
SF03 vs. SF03	SM02 vs. SM02	1.99	0.0480	1
SF03 vs. SF03	SM03 vs. SM03	6.12	1.19E-08	3.34E-07
SM01 vs. SM01	SM02 vs. SM02	-0.0333	0.973	1
SM01 vs. SM01	SM03 vs. SM03	3.05	0.00284	0.0794
SM02 vs. SM02	SM03 vs. SM03	2.01	0.0462	1

Supplemental Table 2.3: Random Forests (10-fold cross validation, 500 trees) label probabilities for each sample for the category individual. Storage temperatures are RT: room temperature (~20 ^{o}C), RF: refrigerated (4 ^{o}C), and FR: frozen (-20 ^{o}C). Values highlighted in yellow correspond to the label with the highest probability that correctly corresponds to the true individual label. Values highlighted in red correspond to the label with the highest probability that does not correspond to the true individual label.

Individual	Storage Time (weeks)	Storage Temperature	Label Probabilities					
			SF01	SF02	SF03	SM01	SM02	SM03
SF01	0	RT	0.678	0.106	0.018	0.008	0.106	0.084
SF01	1	RT	0.778	0.068	0.006	0.016	0.078	0.054
SF01	1	RF	0.75	0.062	0.004	0.014	0.116	0.054
SF01	1	FR	0.746	0.11	0.004	0.018	0.066	0.056
SF01	2	RT	0.734	0.078	0.004	0.01	0.088	0.086
SF01	2	RF	0.722	0.094	0.014	0.022	0.068	0.08
SF01	2	FR	0.782	0.08	0.008	0.004	0.074	0.052
SF01	4	RT	0.726	0.076	0.022	0.026	0.088	0.062
SF01	4	RF	0.59	0.08	0.008	0.02	0.186	0.116
SF01	4	FR	0.708	0.122	0.006	0.022	0.088	0.054
SF01	6	RT	0.652	0.066	0.032	0.028	0.12	0.102
SF01	6	RF	0.746	0.132	0.012	0.018	0.066	0.026
SF01	6	FR	0.74	0.106	0.002	0.014	0.082	0.056
SF02	0	RT	0.128	0.56	0.012	0.114	0.12	0.066
SF02	1	RT	0.208	0.63	0.008	0.03	0.098	0.026
SF02	1	RF	0.332	0.476	0.014	0.034	0.11	0.034
SF02	1	FR	0.086	0.676	0.014	0.048	0.1	0.076
SF02	2	RT	0.208	0.2	0.02	0.056	0.386	0.13
SF02	2	FR	0.14	0.706	0.008	0.024	0.064	0.058
SF02	4	RT	0.27	0.448	0.016	0.052	0.14	0.074
SF02	4	RF	0.242	0.474	0.002	0.09	0.154	0.038
SF02	4	FR	0.064	0.676	0.044	0.076	0.06	0.08
SF02	6	RT	0.224	0.472	0.03	0.144	0.066	0.064
SF02	6	RF	0.078	0.558	0.032	0.064	0.172	0.096
SF03	0	RT	0.008	0.03	0.93	0.01	0.01	0.012
SF03	1	RT	0.04	0.058	0.79	0.036	0.032	0.044
SF03	1	RF	0.018	0.04	0.866	0.036	0.018	0.022
SF03	1	FR	0.034	0.042	0.796	0.058	0.036	0.034
SF03	2	RT	0.02	0.044	0.876	0.04	0.01	0.01
SF03	2	RF	0.012	0.022	0.926	0.014	0.004	0.022
SF03	2	FR	0.032	0.07	0.806	0.048	0.01	0.034
SF03	4	RT	0.024	0.024	0.888	0.036	0.018	0.01
SF03	4	RF	0.024	0.076	0.766	0.08	0.028	0.026
SF03	4	FR	0.038	0.084	0.684	0.088	0.062	0.044
SF03	6	RT	0.03	0.05	0.786	0.058	0.024	0.052
SF03	6	RF	0.042	0.058	0.782	0.038	0.028	0.052
SF03	6	FR	0.052	0.062	0.744	0.054	0.036	0.052
SM01	0	RT	0.036	0.046	0.036	0.712	0.02	0.15
SM01	1	RT	0.034	0.042	0.006	0.85	0.038	0.03
SM01	1	RF	0.022	0.078	0.03	0.78	0.04	0.05
SM01	1	FR	0.052	0.068	0.014	0.702	0.036	0.128
SM01	2	RT	0.03	0.054	0.008	0.788	0.034	0.086
SM01	2	RF	0.018	0.012	0.064	0.78	0.042	0.084
SM01	2	FR	0.038	0.042	0.02	0.694	0.054	0.152
SM01	4	RT	0.034	0.048	0.034	0.694	0.052	0.138
SM01	4	RF	0.032	0.074	0.02	0.662	0.11	0.102
SM01	4	FR	0.026	0.052	0.018	0.802	0.03	0.072
SM01	6	RT	0.054	0.082	0.03	0.712	0.036	0.086
SM01	6	RF	0.044	0.05	0.064	0.672	0.074	0.096
SM01	6	FR	0.036	0.064	0.036	0.596	0.136	0.132
SM02	0	RT	0.07	0.072	0.004	0.008	0.73	0.116
SM02	1	RT	0.044	0.044	0.012	0.04	0.756	0.104
SM02	1	RF	0.05	0.052	0.004	0.026	0.77	0.098
SM02	1	FR	0.088	0.086	0.026	0.028	0.676	0.096
SM02	2	RT	0.186	0.12	0.004	0.03	0.506	0.154
SM02	2	RF	0.36	0.138	0.004	0.04	0.318	0.14
SM02	2	FR	0.154	0.206	0.004	0.016	0.436	0.184
SM02	4	RT	0.08	0.044	0.006	0.022	0.742	0.106
SM02	4	RF	0.138	0.112	0.008	0.01	0.582	0.15
SM02	4	FR	0.086	0.048	0	0.012	0.74	0.114
SM02	6	RT	0.074	0.054	0.004	0.038	0.744	0.086
SM02	6	RF	0.054	0.052	0.002	0.018	0.778	0.096
SM02	6	FR	0.046	0.064	0.012	0.012	0.778	0.088
SM03	1	RT	0.086	0.092	0.036	0.146	0.196	0.444
SM03	1	RF	0.074	0.072	0.012	0.144	0.168	0.53
SM03	1	FR	0.166	0.112	0.02	0.158	0.084	0.46
SM03	2	RT	0.098	0.038	0.03	0.084	0.228	0.522
SM03	2	RF	0.096	0.056	0.01	0.08	0.288	0.47
SM03	4	RT	0.14	0.1	0.002	0.058	0.294	0.406
SM03	4	RF	0.12	0.034	0.014	0.052	0.236	0.544
SM03	4	FR	0.13	0.068	0.024	0.106	0.16	0.512
SM03	6	RF	0.124	0.042	0.002	0.126	0.284	0.422
SM03	6	FR	0.142	0.064	0.03	0.052	0.192	0.52

2.7 References

[1] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2013;486:207-14.

[2] Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. Science. 2009;326:1694-7.

[3] Califf K, Gonzalez A, Knight R, Caporaso JG. The Human Microbiome: Getting Personal. Microbe. 2014;9:410-5.

[4] Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. PNAS. 2010;107:6477-81.

[5] Lax S, Hampton-Marcell JT, Gibbons SM, Colares GB, Smith D, Eisen JA, et al. Forensic analysis of the microbiome of phones and shoes. Microbiome. 2015;3:1-8.

[6] Heath LE, Saunders VA. Assessing the potential of bacterial DNA profiling for forensic soil comparisons. Journal of Forensic Sciences. 2006;51:1062-8.

[7] Horswell J, Cordiner SJ, Maas EW, Martin TM, Sutherland KBW, Speir TW, et al. Forensic comparison of soils by bacterial community profiling. Journal of Forensic Sciences. 2002;47:350-3.

[8] Meyers MS, Foran DR. Spatial and temporal influences on bacterial profiling of forensic soil samples. Journal of Forensic Sciences. 2008;53:652-60.

[9] Damaan FE, Williams DE, Layton AC. Potential use of bacterial community succession in decaying human bone for estimating postmortem interval. Journal of Forensic Sciences. 2015;60:844-50.

[10] Hauther KA, Cobaugh KL, Jantz LM, Sparer TE, DeBruyn JM. Estimating time since death from postmortem human gut microbial communities. Journal of Forensic Sciences. 2015;60:1234-40.

[11] Mayntz-Press KA, Sims LM, Hall A, Ballantyne J. Y-STR Profiling in Extended Interval ($\geq$3 days) Postcoital Cervicovaginal Samples. J of Forensic Sciences. 2008;53:342-8.

[12] Pfeiffer H, Huhne J, Ortmann C, Waterkamp K, Brinkmann B. Mitochondrial DNA typing from human axillary, pubic and head hair shafts - success rates and sequence comparisons. Int J Legal Med. 1999;112:287-90.

[13] Budowle B, Allard MW, Wilson MR, Chakraborty R. Forensics and Mitochondrial DNA: Applications, Debates, and Foundations. Annu Rev Genomics Hum Genet. 2003;4:119-41.

[14] Foran D, Starrs J. In Search of the Boston Strangler: Genetic evidence from the exhumation of Mary Sullivan. Med Sci Law. 2004;44:287-90.

[15] Tridico S, Murray DC, Addison J, Kirkbride KP, Bunce M. Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investigative Genetics. 2014;5.

[16] Lauber CL, Zhou N, Gordon JI, Knight R, Fierer N. Effect of storage conditions on the assessment of bacterial community structure in soil and human-associated samples. FEMS Microbiol Lett. 2010;307:80-6.

[17] Rubin BE, Gibbons SM, Kennedy S, Hapton-Marcell J, Owens S, Gilbert JA. Investigating the Impact of Storage Conditions on Microbial Community Composition in Soil Samples. PLOS ONE. 2013;8:e70460.

[18] Flores R, Shi J, Yu G, Ma B, Ravel J, Goedert JJ, et al. Collection media and delayed freezing effects on microbial composition of human stool. Microbiome. 2015;3.

[19] Gorzelak MA, Gill SK, Tasnim N, Ahmadi-Vand Z, Jay M, Gibson DL. Methods for Improving Human Gut Microbiome Data by Reducing Variability through Sample Processing and Storage of Stool. PLOS ONE. 2015;10:e0134802.

[20] Bai G, Gajer P, Nandy M, Ma B, Yang H, Sakamoto J, et al. Comparison of storage conditions for human vaginal microbiome studies. PLoS ONE. 2012;7:e36934.

[21] Hale VL, Tan CL, Knight R, Amato KR. Effect of preservation method on spider monkey (*Ateles geoffroyi*) fecal microbiota over 8 weeks. Journal of Microbiological Methods. 2015;113:16-26.

[22] Weese JS, Jalali M. Evaluation of the impact of refrigeration on next generation sequencing-based assessment of the canine and feline fecal microbiota. BMC Veterinary Research. 2014;10:230.

[23] Illumina. 16S Metagenomic Sequencing Library Preparation. Part # 15044223 Rev. B ed.

[24] Klindworth A, Pruesse E, Schweer T, Peplles J, Quast C, Horn M, et al. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Research. 2013;41:1-11.

[25] http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/

[26] Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for Illumina sequences. BMC Bioinformatics. 2012;13.

[27] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7:334-6.

[28] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194-200.

[29] Haas BJ, Gevers D, Earl AM, Feldgarden M, Ward DV, Giannoukos G, et al. Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Research. 2011;21:494-504.

[30] Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266-7.

[31] DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol. 2006;72:5069-72.

[32] Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology. 2001;26:32-46.

[33] Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2:18-22.

[34] Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biology. 2010;11:R106.

[35] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15:550.

[36] Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, et al. Characterizing and measuring bias in sequence data. Genome Biology. 2013;14:R51.

[37] Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and function predictions in human microbiome research. PNAS. 2015.

[38] Bartram AK, Lynch MD, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. Generation of Mulitmillion-Sequence 16S rRNA Gene Libraries from Complex

Microbial Communities by Assembling Paired-End Illumina Reads. Appl Environ Microbiol. 2011;77:3846-52.

[39] Quail M, Smith M, Coupland P, Otto TD, Harris S, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:314.

CHAPTER 3

CLASSIFICATION OF INDIVIDUALS AND THE POTENTIAL TO DETECT SEXUAL ACTIVITY USING THE PUBIC MICROBIOME

3.1 Abstract

The pubic hair microbiome has potential in forensic investigations as a tool to distinguish individuals and detect likely sexual contact. Studies characterizing the microbiome of other areas of the body indicate that the microbiome can be individualizing. Here, microbiome profiles were generated from the pubic hairs and swabs taken from the pubic mound region and compared between individuals and couples with the intent to establish the true individualizing nature as well as the ability to detect the transfer of the microbiome associated with the pubic mound. By performing 16S sequencing of pubic hair/pubic area microbiome, samples from 12 couples and 19 matching singles, with varying degrees of sexual activity, I show that a model constructed using a Random Forest classifier predicts which individual samples collected up to six months apart belong to with 90% accuracy. This classifier demonstrates the stability of the microbiome over this time frame. In the context of sexual contact, the pubic hair/pubic area microbiome grouped couples depending on level of sexual activity, but was not able to detect single transfer events. The results establish power of the pubic hair/pubic area

microbiome to identify individuals for forensic associations and associate sexually active couples.

3.2 Introduction

Sexual assaults are a pervasive issue in today's society. According to the World Health Organization, over a third of women experience partner or non-partner sexual violence over their lifetimes [1]. Over 284,000 Americans 12 years of age and older were sexually assaulted or raped in 2014 [2]. In the U.S. Military, there were 6083 victims in reports of sexual assault in fiscal year 2015 [3]. Approximately 30% of rapes are reported with just 1% referred to prosecutors and <1% resulting in conviction [4, 5]. Collection and examination of evidence in cases of sexual assault are very important steps in the resolution of these cases.

While evidence submitted for biological analysis can never answer the question of whether the sexual contact was consensual, the evidence can be screened for bodily fluids and DNA to support the claims that the contact occurred. Detection of semen on intimate swabs (vaginal, cervical, and/or rectal swabs) by chemical, immunological, and microscopic means followed by DNA analysis is currently the definitive diagnostic approach to demonstrating that sexual contact has occurred. When no semen is found on an intimate swab, additional collected evidence such as other body swabs, underwear, clothing, or bedding may be examined for the

presence of semen. As examinations move away from the intimate samples it becomes harder to show that physical sexual contact has occurred. Nevertheless, residual evidence such as pubic hairs may provide evidence of presence, and here I reason that microbiome analysis of hairs or skin in the pubic region might provide evidence of actual sexual contact. Currently, pubic hair evidence is not commonly tested in crime laboratories, as it must first undergo microscopic evaluation, an uncommon test due to lack of trained examiners and lack of scientifically validated methodology [6]. The rationale for this study is to establish whether the pubic hair microbiome has the potential to assist investigators of sexual assault when other methods do not produce results.

Microbiome studies are increasingly showing the importance of the microbiome in human health [7] and the environment [8-10]. Practical applications of this knowledge through interventional therapies have had varying success [11, 12]. The National Institutes of Health (NIH) Human Microbiome Project (HMP) was launched in 2007 with the goals of characterizing the human microbiome and showing opportunities for improvement for human health through the monitoring or manipulation of the human microbiome [13]. From this study and others [14-16], variation between individuals has been shown to be larger than within an individual at a given body site and that the microbiome may be individualizing, characteristics which may be useful for forensic purposes.

While much microbiome research focuses on the function of the microbiome and its relationship to human health, it is also of interest to test the extent to which microbiome taxonomic composition can be used to distinguish individuals in forensic investigations. The ability to discriminate between individuals and detect the transfer of the microbiome between individuals may be especially useful when other more-common testing methods fail to produce useful results. There is some evidence that bacteria are exchanged to varying degrees during intimate contact, more casual contact, or simply by sharing of the living environment. When inoculated with saliva, the length of time it took for various skin sites to return to their baseline microbiome varied between dry versus sebaceous skin sites [15]. In the case of oral microbiome exchange [17], one participant was inoculated with bacteria through a probiotic drink prior to kissing. At each exchange (drink to participant one, participant one to participant two), it became increasingly difficult to detect the inoculant even with the spike-in from the probiotic drink. It has also been shown that the microbiome of co-habiting family members, including dogs, are more similar to each other on hand/paw surfaces than to unrelated individuals [18]. Given the variability in the extent of microbiome transfer and the ability to detect such transfer, it is important to model as best as possible the scenario under which one wishes to use the microbiome for forensics. Specifically, before adding the microbiome to the investigator's toolbox, one must first establish that the microbiome can be used reliably to differentiate individuals and, if this is possible, determine the timeframe over which such transfer events are detectable.

To date, the use of the microbiome, whether it be human or otherwise, has seen limited application in forensics [19]. A notable exception is the use of soil microbiome for geolocation, time of death estimates, or touch DNA [20]. Using a mouse model and samples collected from both the mice and the soil upon which the mice rested, Metcalf et al. provide proof of concept for a microbial clock for estimating the postmortem interval [21]. Additional studies on the metagenome during cadaver decomposition [22, 23] and changes to soil metagenome composition during cadaver decomposition [24, 25] add to the growing body of work on this topic. While there have been studies on the detection of individuals on surfaces using the microbiome [26] and the ability to detect the microbiome on surfaces over time [27], studies addressing the ability to detect the transfer of the microbiome directly between individuals are much more limited.

The human hair microbiome has received little attention, possibly because there are few obvious health implications associated with hair-associated microbes. Tridico et al. [28] evaluated the microbiome of human scalp and pubic hair for potential forensic use and provided the first indication that the pubic hair microbiome may be useful for detection of sexual contact. There was temporal stability in the “core” pubic hair microbiome but conclusions were limited due to the small sample size of seven individuals. We also showed in Chapter 2 that pubic hairs sampled at a single time point from six subjects and stored under various temperatures and for various timeframes [29] allowed individuals to be distinguished from one another across the storage conditions.

Here I quantify the ability to separate individuals based on their microbiomes with the potential to re-identify an individual with samples collected over a period of months. Surveys of participant hygiene and sexual activities provide preliminary insight into how these activities may affect the microbiome recovered from an individual. Additionally, results from the couples suggest some commonalities within the couples, and comparison with results from other skin sites allow inferences concerning the potential to detect transfer of microbiome during sexual contact.

3.3 Methods

3.3.1 Sample Collection

Adults who self-identified as being free of sexually transmitted disease (STD) and yeast infection at time of enrollment were recruited from the Georgia Institute of Technology and Defense Forensic Science Center (DFSC). Volunteers who were sexually active were requested to ask their sexual partner to also participate in the study. At time of enrollment, five collection packets coded with a unique alphanumeric identifier were provided to the participant to protect their privacy and ensure accuracy of sample tracking. The coded collection packets contained a sterile suture removal kit containing scissors and forceps, sterile plastic tongs,

sterile 50 mL urine collection vials, two sterile cotton-tipped swabs, one cardboard swab box, one disposable dropper vial of sterile water, and collection instructions. Participants were instructed to collect pubic mound hair and pubic mound swab samples at three set time points: week 0, week 6, and week 12. If the participant did not have pubic mound hair, they were instructed to collect the pubic mound swabs only. The sexually active couples were asked to collect at least two additional samples at approximately two-week intervals from another collection time point, where possible within 72 hours of sexual activity. For each collection, participants filled out an online survey (Supplemental Figure 3.1). The participants were instructed to collect at least 10 cut pubic hairs from all around the pubic mound area with the females avoiding hairs from the labial area or hairs in direct contact with the vaginal opening.

In a separate collection conducted 6 months later, a random subset of 20 participants, including 5 of the 12 pairs of couple participants, was requested to collect swab samples from each inner elbow and behind each ear. Participants were provided with two collection packets, each containing eight sterile cotton-tipped swabs, four labeled cardboard swab boxes, and two disposable dropper vials of sterile water. For each collection, spaced two weeks apart, the participant was requested to collect two swabs each of the right inner elbow, the left inner elbow, behind the right ear, and behind the left ear. No survey was associated with these sample collections.

For all collections, the completed collection packets were held at room temperature until the participant returned the packet to the provided drop box. Median time to drop-off was one day, within the period shown in Chapter 2 not to adversely affect sample preservation for the purposes of identity matching. After delivery, collection packets were stored at -4 ^{o}C until samples were processed for DNA extraction.

3.3.2 DNA Extraction and Sequencing

At least 7 cm of total combined hair length, when possible, was used for each extraction with hair extractions performed in duplicate. Hairs were cut to be approximately 1 cm in length before being put into the extraction tube. Where there was not enough hair sample to perform two extractions with at least 7 cm of total hair length in each, the hair was roughly divided between the two extractions. The head of each swab collected at each pubic area time point was removed for DNA extraction. When only one swab was provided, either the single swab was extracted without a duplicate or the swab was divided in half for replicate extractions. For the elbow and ear swabs, the right elbow and ear swabs were used for analysis. Each body area was extracted in duplicate with half of each swab split between the two sample tubes. Genomic DNA was extracted using the MO BIO/QIAGEN PowerFecal® DNA Isolation Kit following manufacturer's directions with the modification of performing the final elution with 50 μl Solution C6. Extractions were performed at each collection time point with a reagent blank included with each extraction set. Extracts were concentrated to dryness using a

vacuum concentrator and stored at -20 ^{o}C until library preparation. Prior to targeted amplification, samples were reconstituted in 12 μl amplification-grade water.

Sequencing libraries were prepared following the Illumina 16S Metagenomic Sequencing Library Preparation workflow [30] with modifications as previously described [29] with 300 bp paired end sequencing on a MiSeq at DFSC, Atlanta, GA. Amplification of the V3 and V4 region of the 16S rRNA gene was performed with 2.5 μl of reconstituted sample and purified amplicons were normalized to 50 pM. Each pooled library contained 40-50 samples. Primary sequencing was performed in 17 sequencing lanes. Samples were grouped in sequencing runs based on order received so that samples from an individual were spread across multiple sequencing runs. An additional sequencing run each was used for the re-sampled hairs and ear/elbow swabs.

3.3.3 Sequence Analysis and Community Comparisons

Prior to community analyses, a series of quality filtering steps were performed on the sequences. Sequences with a base quality score (Q) of less than 25 and a sequence length of less than 100 base pairs (bp) were removed with Trim Galore! [31] and forward and reverse reads were merged with the PANDAseq Assembler [32]. Closed reference OTU picking (*pick_closed_reference_otus.py*) was conducted using QIIME 1 [33], during which sequences were clustered based on 97% similarity, using the default settings (UCLUST [34], PyNAST alignment [35],

Greengenes 13_8 [36]). A separate chimera-removing step was not employed as the Greengenes database has been curated to remove chimeras. Any sequences not aligning to the Greengenes database were removed. Samples with fewer than 10,000 reads were re-amplified and sequenced as above in a single sequencing run. The samples with the highest read counts between sequencing efforts were retained. Since Rhizobiales was found to be predominant in the reagent and amplification blanks, but is known to not be a skin commensal microorganism, this genus was filtered from the OTU tables. OTUs below a total abundance of 0.01% were filtered from the combined OTU table to minimize stochastic effects. To avoid artifacts due to variable read depth, random rarefaction to 4428 reads per sample was performed using the *core_diversity_analysis.py* script. This level balances maximizing reads with minimizing loss of information for generation of taxonomy plots, UniFrac distances, and PCoA plots.

Alpha diversity plots were generated (*alpha_rarefaction.*py) to determine a rarefaction level to use for beta diversity analysis. Alpha diversity was used (*alpha_diversity.py*, *compare_alpha_diversity.py*) to evaluate the variation in alpha diversity over various categories with a non-parametric (Monte Carlo) two-sided Student's two-sample t-test (Bonferroni corrected) with 999. Beta diversity comparisons were performed with the vegan::adonis nonparametric method for multivariate analysis of variance using the weighted and unweighted UniFrac distances (*compare_categories.py –method adonis*) [37].

3.3.4 Deblur Sample Processing

The Deblur denoising algorithm [38], using the QIIME 2 framework, was performed on the forward reads only to resolve finer differences between sequences. The raw sequence files were trimmed to 301 bp and a minimum Q score of 5 using Trim Galore!. Deblur was then conducted using a trim length of 234 bp, chosen by determining where the sequences tended to fall below Q20, and default settings. Taxonomic classification was performed using the QIIME 2 feature classifier trained to the Greengenes 18_7 99% OTU database with the primers as described before [29] and 300 bp truncation length. Core diversity analysis (alpha and beta diversity) with PERMANOVA beta group significance was conducted with the QIIME 2 diversity plug-in.

3.3.5 Supervised Learning Classification

Classification of individuals and couples was performed using Random Forests [39, 40]. Initial 10-fold cross-validation was performed using the implementation of Random Forests in QIIME 1 (*supervised_learning.py -e cv10*) with default number of trees (500). This method was used to determine the classification error rate for the categories ‘Individual’ and ‘Couple.’ Samples were analyzed using the full pubic region data set, just the swab samples, just the hair samples, the elbow samples, or the ear samples.

In order to assess the ability of the Random Forest model to predict the ‘Individual’ label, out-of-bag (OOB) Random Forest calculations fitting the model

to the category Individual using the full pubic area data set (hairs and swabs) and 500 trees were also performed using the direct implementation of the 'randomForest' package in R. The resulting randomForest object was used to predict the 'Individual' label of a subset of hair samples re-sampled from the original collections and given a new sample designation for blind testing and samples previously collected to test storage conditions [29]. The re-sampled hairs were processed in triplicate as above. Based on incidental information provided at the time of collection of the storage samples, it was known that two individuals, a male and a female, participated in both the storage study and the current study, however, which samples from the storage study belong to these individuals was not ascertainable due to IRB restrictions. Additionally, the remaining two females from the storage study were known not to have participated in the current study. It was unknown whether the remaining two males from the storage study participated in the current study. As the storage data was originally processed with open reference picking, raw sequencing files were re-processed as above from the Trim Galore! step (trim length 100, Q20) with closed reference picking and filtering as described above. Samples with fewer than 900 reads were removed. In order for the storage data to be compatible with the current study data, any OTUs present in the current data but not present in the storage data needed to be added to the storage OTU table with abundances of '0' for these OTUs. Predictions were performed using the predict function of the R 'stats' package with the randomForest object. For the re-sampled hairs, predictions were made on each replicate separately and by collapsing the replicates prior to

predictions. After predictions were made on the re-sampled hairs the predicted 'Individual' labels were compared to the original 'Individual' labels.

3.3.6 Hierarchical Clustering

Hierarchical clustering was also employed as an alternative approach to assess the ability to differentiate between couples. A presence/absence table of OTUs was created by first collapsing the read data using QIIME 1 by summing the reads for each participant by collection week and hair or swab or hair and swab jointly (*collapse_samples.py*). That is to say, the two or three independent samples for each pubic region were pooled for each individual at each time point, so as to reduce stochastic sampling noise. The resulting table was then converted to presence/absence (1/0) in Microsoft® Excel. Each individual was assigned a couple designation. If a pair of individuals were part of an actual couple, they were given the same couple designation. Otherwise, an individual was given a unique couple designation. Hierarchical clustering using the Ward Method with no standardization of the data was conducted in JMP® (SAS Institute Inc.). Clustering was performed with 7-15 clusters on the hair and swabs samples together, hair samples only, and swab samples only. Using R [41], the couple designations were randomly shuffled for 10,000 permutations. Pearson's Chi-square Test (*chisq.test*) was conducted on each contingency table formed from couple and cluster assignments. Histogram tables for each cluster variation were generated. For each cluster, the Chi-square statistics for the permutations and true couple assignment were ordered largest to smallest. A P-value was

generated by taking the rank of the true couple assignment and dividing by 10,001 (number of permutations plus the true couple assignment). For comparison, clustering was repeated for the Deblur-processed samples at 12 and 15 clusters using the hair and swabs together.

3.3.7 Closest Non-self Neighbor Comparison

A custom script was created in R that, for each sample, used the weighted or unweighted UniFrac distance matrix to determine which non-self sample was the nearest in order to investigate whether couples were more related using this metric. It was then determined how many samples for each couple had their partner as the nearest neighbor. This analysis was conducted on the pubic hair/swab samples, the elbow samples, and the ear samples.

3.3.8 Comparison to Other Skin Sites and Buccal Mucosa

Sample taxa were compared to reported skin taxa from Grice et al. [42] in order to determine to which skin area the pubic hair/pubic area microbiome is most similar. The pubic hair/pubic area samples were collapsed as before by ‘Individual’ and the reads were summed. The genera that corresponded to those in [42] were further collapsed within an individual and summed. Relative abundances were calculated for all samples, hair and skin, and used for Principal Components on Correlation analysis in JMP® with the default estimation method.

In order to investigate the potential to detect oral contact to a participant's pubic region, a random subset of 40 16S trimmed fasta sequences (20 male, 20 female; Supplemental Table 3.1) from buccal mucosal samples from the Human Microbiome Project [14, 43] were used to perform closed reference OTU picking. The beta diversity of these sequences was compared to the beta diversity of the pubic hair and pubic swabs using weighted and unweighted UniFrac at 4428 reads and visualized through PCoA plots. The use of the 4428 read rarefaction level resulted in 10 of the buccal mucosal samples dropping from the analysis. The average UniFrac distance, weighted and unweighted, of each pubic area hair/swab sample to each HMP buccal mucosal sample was calculated with the no oral contact and oral contact groups compared using a two-sample t-test.

3.4 Results

3.4.1 Demographic Summary

A total of 43 individuals, including 12 partner pairs, provided pubic hair and/or pubic mound area swabs. Participants provided 1 to 5 sets of sample collections (hairs and/or swabs at a single time point) resulting in 155 completed sample collections, an average of 3.6 collections per person. Due to the use of various hair removal techniques employed by some participants, more sets of swabs were collected than hairs (148/55% and 123/45%, respectively). The demographic breakdown and summarized survey results are presented in Table 3.1. The ages

Table 3.1 Summarized demographic and survey responses of participants.

Gender*			
	Male	23	53.5%
	Female	20	46.5%
Racial Group*			
	Caucasian non-Hispanic	37	86.0%
	Caucasian Hispanic	4	9.3%
	Southeast Asian	1	2.3%
	No response	1	2.3%
Age*			
	21-30	7	16.3%
	31-40	16	37.2%
	41-50	8	18.6%
	51-60	7	16.3%
	61-70	5	11.6%
Bathing Frequency**			
	Once a day	130	83.9%
	Every other day	13	8.4%
	Multiple times a day	9	5.8%
	No response	3	1.9%
Antibacterial Soap Usage**			
	No	120	77.4%
	Yes	11	7.1%
	Unknown by participant	21	13.5%
	No response	3	1.9%
Hair Removal**			
	None	86	55.5%
	Yes		
	Shave	57	36.8%
	Laser	8	5.2%
	Wax	1	0.6%
	No response	3	1.9%

Antibiotic Usage Within Month of Collection**			
	No	137	88.4%
	Yes	15	9.7%
Sexually Active**			
	Yes	122	78.7%
	No	30	19.4%
	No response	3	1.9%
Sexually Active in 7 Days Prior to Collection**			
	Yes	84	54.2%
	No	67	43.2%
	No response	4	2.6%
Frequency of Sexual Activity in 7 Days Prior to Collection***			
	1	37	44.0%
	2	33	39.3%
	3	8	9.5%
	4	5	6.0%
	5	1	1.2%
Condom Use During Prior Sexual Activity***			
	No	59	70.2%
	Yes	23	27.4%
	No response	2	2.4%
Partner Oral Contact to Genitals***			
	No	58	69.0%
	Yes	22	26.2%
	No response	4	4.8%

* Out of 43 Participants
** Out of 155 total sample collections
*** Out of 35 participants sexually active in 7 days prior to collection/84 sample collections

of the participants ranged from 25 to 68 years of age. Extracted hair lengths ranged from 1.4 cm to 12 cm (mean 7.9 cm, median 7.8 cm).

Within the group of participants who indicated that they were sexually active, participants varied in their level of sexual activity in the seven days prior to sample collection. Ten of 12 couples and 10 individuals without partner participation in the study reported being sexually active prior to sample collection. In the seven days

prior to sample collection, three partner pairs reported having sex in the prior week for all of their sample collections and seven partner pairs were mixed in their level of sexual activity prior to their sample collections. Half of the sexually active individuals without partner participation in the study also reported sexual activity prior to all collections. The two couples who reported not being sexually active prior to all sample collections were in their low-to-mid forties. The average number of sexual encounters in the week prior to collection was 1.8. There was a high concordance in couples self-reporting sexual activity during the collection week with mismatches in responses occurring when the partners performed sample collections on different days during the same week. Condom usage and oral contact to genitals by sample collection was 70.2% and 26.2%, respectively. There was just a single case of more than one partner and no self-reported homosexual activity.

3.4.2 Sequencing Summary

After quality trimming and filtering for OTUs below 0.01% total abundance, there were at total of 113,682,691 reads over 536 samples, ranging from 1335 reads to 6,949,912 reads per sample. The median read count was 120,235 reads. This data set represents 307 OTUs with 91 unique genus/species (Figure 3.1). Eleven samples were below 4428 reads and eliminated from further analyses at that rarefaction level. Similar to previously reported skin microbiome results [42], the predominant bacterial phyla across all samples were Actinobacteria (49.1%), Firmicutes (46.6%), and Proteobacteria (2.9%). Over 77% of the genera were

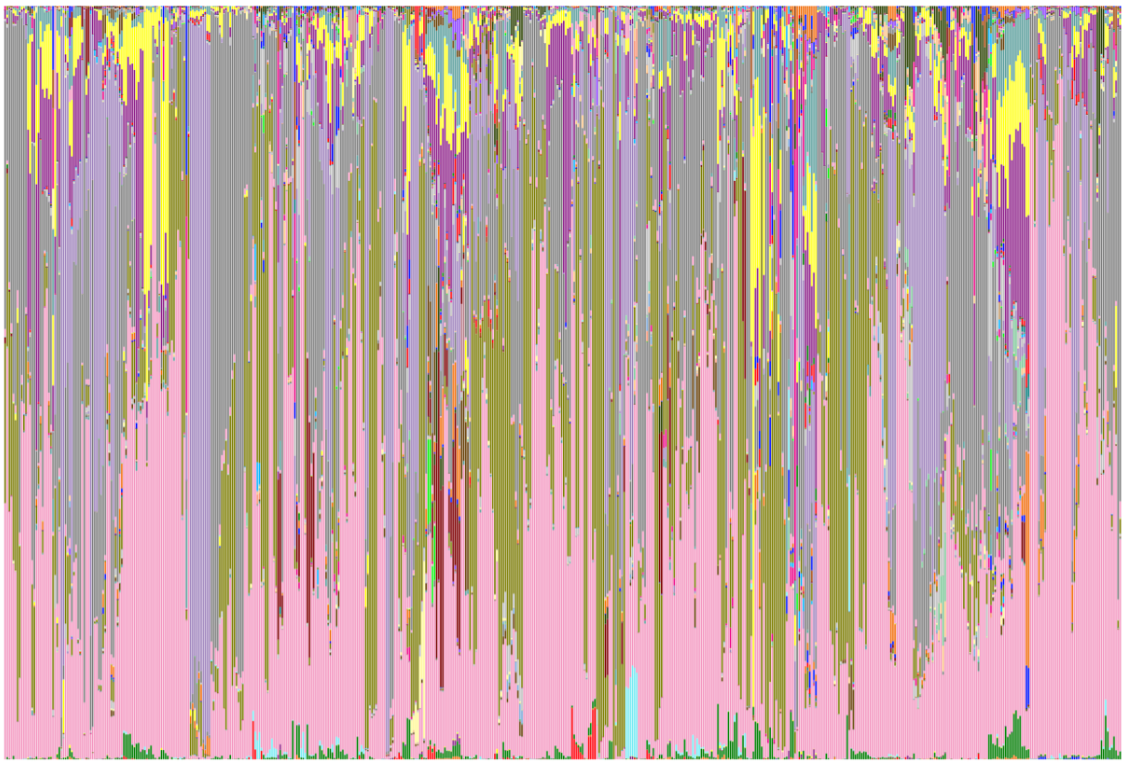


Figure 3.1 Genus level taxonomic bar chart of pubic area microbiome samples. Samples are sorted by individuals then time point. Taxa below 0.01% overall abundance have been filtered from the samples.

represented by *Corynebacteria* (29.2%), *Staphylococci* (21.5%), *Propionibacteria* (15.4%), and *Lactobacilli* (11.5%). Comparing the pubic area microbiome taxonomy to that of sebaceous, moist, and dry skin sites from [42], the pubic area microbiome is more similar to that of sebaceous skin sites than to dry sites (Supplemental Figure 3.2). Differentially abundant genus-level OTUs were present between men and women, with 34 *Corynebacterium*-associated OTUs (out of 109 OTUs enriched in males) more abundant in males and 23 *Lactobacillus*-associated OTUs (out of 95 OTUs enriched in females) more abundant in females. The remaining demographic classifications, race and age,

were not sufficiently populated to be able to make meaningful comparisons of differentially abundant OTUs.

3.4.3 Alpha and Beta Diversity

Alpha diversity varied based on sample type (hair or swab) and participant hygiene and sexual practices (Figure 3.2). These trends were similar across the various metrics (number of OTUs, Faith's Phylogenetic Diversity, and Chao1; data not shown). Females had higher alpha diversity than males (p=0.002). The microbiome from antibacterial soap users was less diverse than that from those who do not use antibacterial soap. And those who bathed more frequently (multiple times a day) were less diverse than those who bathed every other day or once a day. Those who had sex in the seven days prior to sample collection had a higher alpha diversity than those who had not (p=0.003). Swabs had higher alpha diversity compared to hairs (p=0.001), but the two sample types from the same individual tended to be highly similar to one another. The clustering between hairs and swabs, as shown by Figure 3.2, may demonstrate the differences in sample collection technique where the swabs are collecting skin microbes in addition to hair microbes or may represent a difference in the microbial richness between hair and the skin. Alpha diversity differences between antibiotic users/antibiotic non-users and those with oral contact/those who did not have oral contact (who were sexually active in prior week) were not significant (p>0.05). Those who used a hair removal technique had higher alpha diversity than those who did not, but this

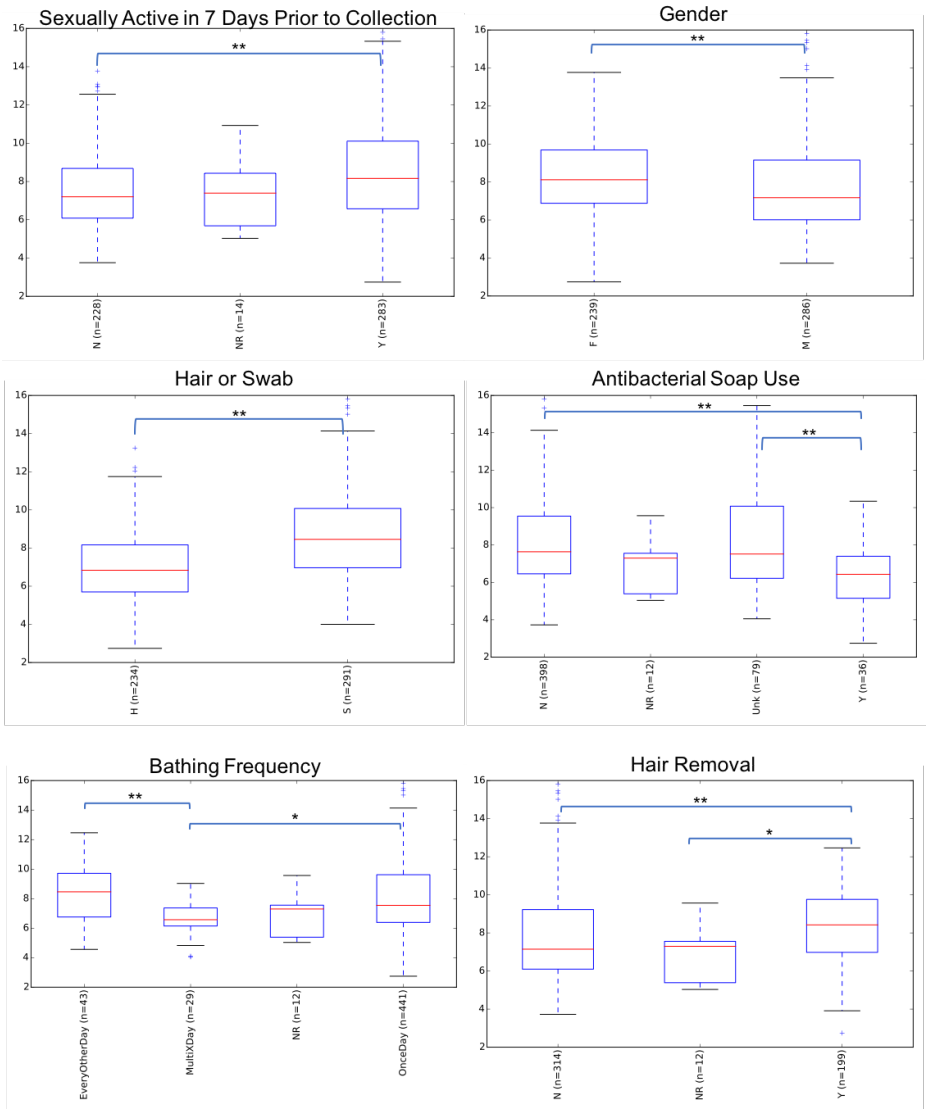


Figure 3.2 Comparisons of Faith's Phylogenetic Diversity (y-axis) for various metadata categories. The number of samples in each category is represented in the parentheses. Red lines represent the median alpha diversity value. The boxes represent the first through third interquartile range. Minimums and maximums are represented by the whiskers with the outliers indicated by +. * $0.01 < p < 0.05$, ** $0.001 < p < 0.01$.

result may be confounded by the predominance of swabs, which have higher alpha diversity than hairs, for these samples. Condom users (81 samples) had higher alpha diversity than those who did not use condoms (194 samples)

(p=0.006). The alpha diversity differences in extracted hair length, binned in 1 cm increments from less than 3 cm to 12 cm, was not significant between any two bins (p>0.05).

Principal components analysis based on the weighted UniFrac metric revealed sample clustering based on 'Individual' and 'Gender' (Figure 3.3 a, c), but there was no clear visual clustering based on 'Couple' (Figure 3.3 b,e). Given that there are a few OTUs within a sample that represent most of the reads, it was theorized that the unweighted UniFrac metric, which is more sensitive to the rarer OTUs than the weighted UniFrac metric is, would exhibit more clustering if couples tended to share more of their rarer OTUs than unrelated individuals. However, clustering patterns did not differ between unweighted and weighted analyses (Figures 3.3 d-f). Using the adonis method for analyzing the statistical significance of the categories, 'Individual' was responsible for the largest effect size (weighted UniFrac R^2=0.522, p<0.001; unweighted UniFrac R^2=0.401, p<0.001) with 'Couple' providing the next largest effect size (weighted UniFrac R^2=0.379, p<0.001; unweighted UniFrac R^2=0.319, p<0.001) (Table 3.2). Overall, the weighted UniFrac distances produced larger effect sizes than the unweighted UniFrac distances. Using Deblur processed data there were no individuals who were solely significantly associated with themselves or their partner using either the weighted or unweighted UniFrac (data not shown). While there was a difference in alpha diversity between hairs and swabs, this category was not significant overall in the beta diversity but retained separation within an individual.

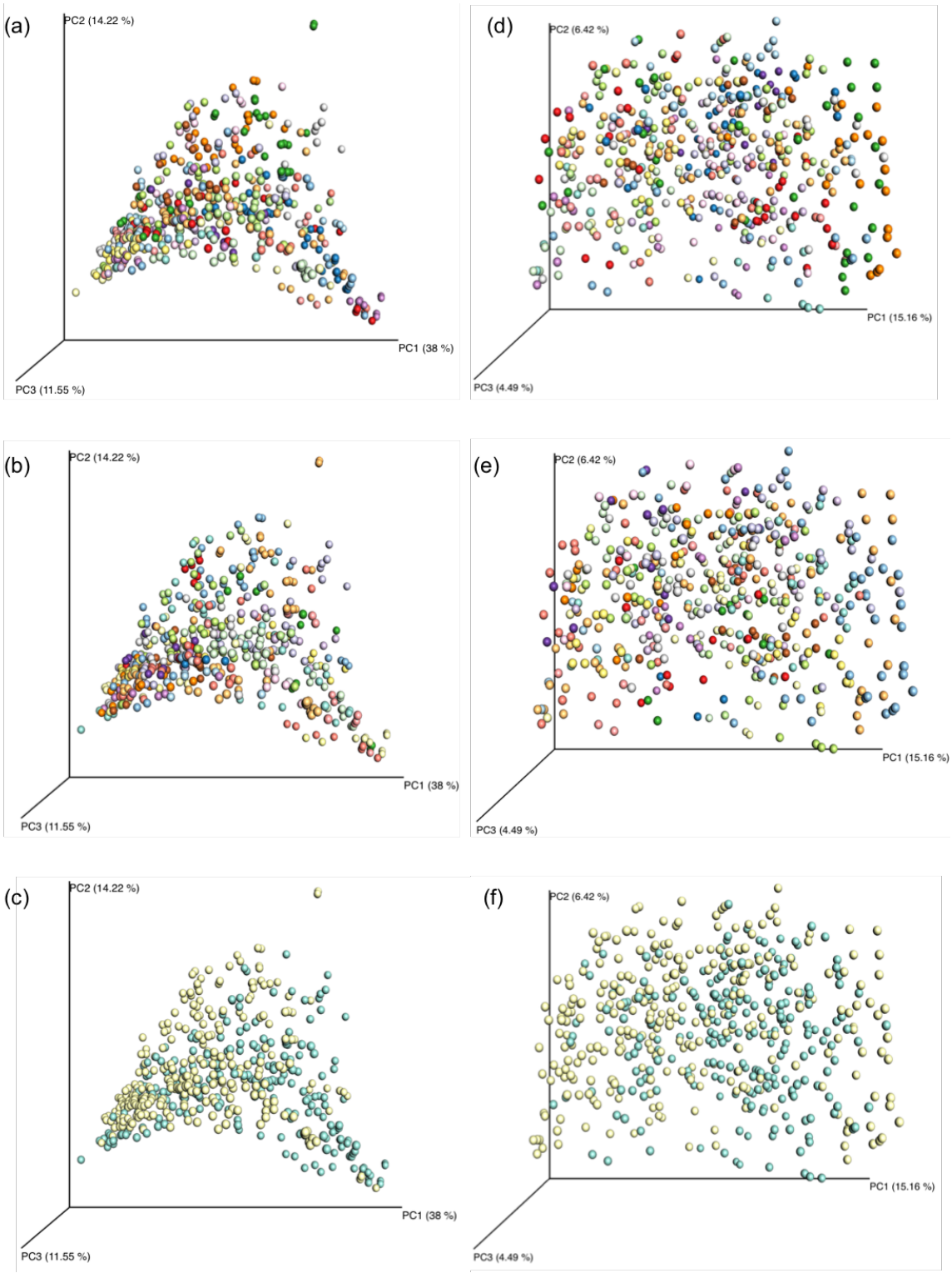


Figure 3.3 Principal components plots of the weighted (a-c) and unweighted (d-f) UniFrac distances for 'Individual' (a, d), 'Couple' (b, e), and Gender (c, f).

Table 3.2 Effect sizes of various metadata categories with the vegan::adonis method based on weighted and unweighted UniFrac distances (hairs and swabs together).

		Weighted UniFrac			Unweighted UniFrac		
Category	df	F	R^2	P-value	F	R^2	P-value
Individual	42	12.5	0.522	<0.001	7.67	0.401	<0.001
Couple	30	10.0	0.379	<0.001	7.71	0.319	<0.001
Gender	1	38.7	0.069	<0.001	16.8	0.031	<0.001
Sex Prior Week	2	7.10	0.026	<0.001	4.43	0.017	<0.001
Antibacterial Soap Use	3	5.67	0.032	<0.001	5.67	0.018	<0.001
Antibiotic Use	2	6.05	0.023	<0.001	6.05	0.014	<0.001
Bathing Frequency	3	9.13	0.050	<0.001	9.13	0.021	<0.001
Hair or Swab	1	8.32	0.016	<0.001	12.8	0.024	<0.001
Extracted Hair Length (binned)	9	2.24	0.038	<0.001	2.64	0.044	<0.001

F: F statistic
df: degrees of freedom
R^2: effect size

In order to attempt to detect transfer of salivary bacteria during oral sex, HMP buccal mucosal samples were compared with the pubic hair and swab samples. Both the weighted and unweighted UniFrac PCoA show a separation between the two sample types, with a greater separation based on unweighted distances. In the weighted UniFrac, most of the separation between the buccal mucosal samples and the pubic samples was along PC3 (PC1 vs PC3 and PC2 vs PC3) with a small amount of overlap with the pubic hair and swab samples. There is no obvious visual grouping difference between the oral contact samples and no oral contact samples with either distance metric. The differences in average distances between the pubic area hair/swab samples and the HMP buccal mucosal samples were not significant between the oral contact and no oral contact groups for either the weighted UniFrac (p=0.245) or the unweighted UniFrac (p=0.065) distances.

3.4.4 Random Forests Modeling

Individualization and couple-sharing were assessed by Random Forest classification with 10-fold cross-validation. This analysis performed significantly better than random guessing for both Individual and Couple classifications for the pubic swabs/hairs (Table 3.3), with typical misclassification rates less than 10%. This was true whether the classification was conducted using all individuals, or

Table 3.3 Random Forest 10-fold cross-validation estimated error based on 500 trees.

	Estimated Error			Ratio Baseline Error
	Mean	Standard Deviation	Baseline Error	to Observed Error
Individual				
Pubic Swabs & Hairs	0.10	0.04	0.96	9.43
Pubic Swabs Only	0.06	0.04	0.97	17.50
Pubic Hairs Only	0.12	0.08	0.96	7.78
Elbow	0.07	0.05	0.95	14.40
Ear	0.01	0.03	0.95	72.00
Couple (All Individuals)				
Pubic Swabs & Hairs	0.08	0.05	0.93	11.47
Pubic Swabs Only	0.06	0.03	0.93	14.79
Pubic Hairs Only	0.15	0.07	0.92	5.96
Elbow	0.03	0.05	0.89	33.40
Ear	0.03	0.05	0.89	33.40
Couple (Couples Only)				
Pubic Swabs & Hairs	0.05	0.04	0.87	17.42
Pubic Swabs & Hairs*	0.05	0.09	0.73	13.37
Pubic Swabs Only	0.05	0.05	0.88	19.49
Pubic Swabs Only*	0.02	0.05	0.70	42.09
Pubic Hairs Only	0.14	0.11	0.85	6.20
Pubic Hairs Only*	0.13	0.12	0.68	5.26
Elbow	0.00	0.00	0.82	N/A
Ear	0.17	0.17	0.58	3.45

N/A = Not Applicable
* Only couple samples corresponding to elbow/ear couples

using just couples to determine the impact of individuals on the couple classification. Accuracy was assessed by evaluating the ratio of the mean baseline classification error to the estimated classification error, where a high ratio means that the individual or couple classification is correct much more often than not (that is, rarely in error). For the pubic area, the classifier performed best when using only the swabs, followed by hairs and swabs together, and then hairs only.

In order to evaluate the role of contact versus shared environment on the ability to correctly classify couples, the classification of samples collected from the inner elbow and behind the ear were also evaluated using Random Forest. It was thought that the skin behind the ear would have less direct transfer between individuals and would be less similar between individuals comprising a couple. When using all of the samples together to build the Random Forest model per body area, the elbow and ear areas showed levels of individualization and couple classification similar to the pubic area (Table 3.3). When using just the elbow and ear samples from the couples to build the Random Forest model, all of the couples were correctly classified for the elbow samples, hence a mean error of zero with no standard deviation, and the mean error of the ear classifications increased. Limiting the pubic hair and swab samples just to those couples with corresponding elbow and ear samples resulted in similar error rates to when all of the couples were used in the Random Forest model. This result implies that

couples share microbiome profiles from these supposedly hidden sites at least to the same degree as they share pubic mound profiles, casting doubt on the necessity of sexual contact to generate the pubic mound similarities.

The Random Forest model was further validated by resampling 10 from the original hair collections and predicting the identity of the newly sequenced samples by generating predictions for each replicate separately and by pooling the replicates. The correct identity was predicted for 9 out of 10 newly processed samples. For these 9 samples, the correct prediction was achieved both when predicting each replicate separately and when pooling the replicates. For the sample that had the wrong identity prediction, one of three of the replicates predicted the correct identity while the remaining replicates and the pooled data were incorrect. The gender of all of the incorrect predictions matched the correct gender of the individual.

Of further interest was how the Random Forest model created from this data would perform when used to predict the classification of an individual in a previously generated storage study data set [29] that contained some common individuals. The overall out-of-bag estimate of error rate in the training set of data, containing the OTUs of the hairs and swabs, was 10.8%. Using the same Random Forest model trained on individual, there were two individuals in the storage study data set whose samples each matched to single individuals (Table

3.4). All eight samples from female individual SF02 were classified as individual PF02. All thirteen samples from male SM02 were classified as individual PM18. While the identities of individuals SF02 and SM02 are unknown, individuals PF02 and PM18 are known to have contributed samples to both studies. Storage study samples were collected in August 2015 while current samples were collected in April-July 2016 for SM02 and July-October 2016 for SF02, an interval of 8-14 months between collections. For the remaining two female participants from the storage study, each with 13 samples and known to have not participated in the current study, there were no consensus individual classifications. For the

Table 3.4 Predicted individual classification of storage study samples. Blue - all samples predicted to be one individual. Green - prediction matches correct gender of individual. Red - prediction does not match gender of individual. Individuals with no corresponding predictions not included.

Predicted Classification	Original Classification					
	SF01	SF02	SF03	SM01	SM02	SM03
PM23	3	0	0	0	0	0
PM94	0	0	0	1	0	2
PM18	0	0	0	0	13	0
PM41	2	0	0	1	0	2
PF21	0	0	3	0	0	0
PM14	5	0	0	3	0	0
PM31	0	0	6	0	0	0
PM73	2	0	0	4	0	0
PF02	1	8	0	0	0	0
PF93	0	0	2	0	0	0
PM24	0	0	0	4	0	4
PM44	0	0	2	0	0	0
PM66	0	0	0	0	0	1
Total samples	13	8	13	13	13	9

remaining two males from the storage study, with 13 and 9 samples and unknown participation in the current study, there were again no consensus individual classifications. These data confirm the high potential for pubic mound microbiome samples to provide consistent signatures of an individual over a period of 18 months.

3.4.5 Hierarchical Clustering

Hierarchical clustering on the presence/absence matrix for each OTU in each sample was used to explore the intra- and inter-couple relationships given there was no clear clustering using UniFrac beta diversity measures and PCoA. No couples shared a cluster exclusive of other couples or individuals, but there is evidence that some couples are more similar than expected by chance. For example, as shown in Figure 3.4, almost all of the samples for couples 5, 9, and 12 cluster together. Some couples cluster only at certain time points, such as couple 11. It is noteworthy that Couples 7 and 8, who did not have sex in the seven days prior to sample collection for any of the time points, did not share any clusters within the couples, and they were the only couples who did not. For the remaining couples, there was no consistent pattern for when the individuals in a couple would be in the same cluster compared to when they were not. Repeating the clustering using the Deblur-processed OTUs with hair and swab data fifteen clusters did not improve the couple clustering overall (data not shown).

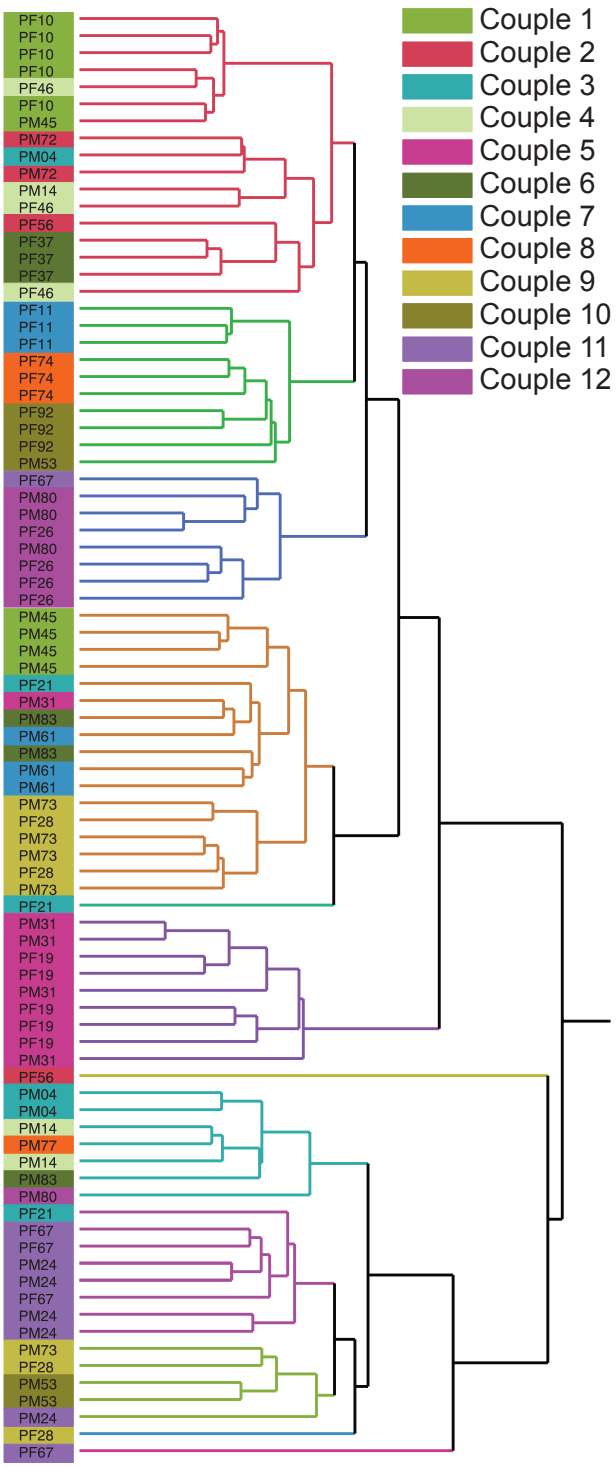


Figure 3.4 Grouping of couples in the hierarchical clustering when collapsed by individual and time point. Samples are grouped into 12 clusters. Only samples from couples and swabs are shown. Note the high degree of co-clustering for couples 5, 9, and 12; and the absence of clustering for couples 3, 6, and 7.

The significance of the clustering evaluated by performing 10,000 iterations of random shuffling of the couple assignments when using hairs and swabs and swabs only (Table 3.5, Supplemental Figures 3.3-3.5). For swabs only, the observed data gave the highest degree of couples clustering across the range of 7 to 15 clusters (permutation $p=10^{-4}$), but for one iteration with 10 clusters only,

Table 3.5 Comparison of Chi Square (couple by cluster) of actual couple assignment and random permutations.

Number of Clusters	Hairs and Swabs				Swabs Only				Hairs Only			
	Mean ChiSq	True ChiSq	Std. Dev. Difference	Rank P	Mean ChiSq	True ChiSq	Std. Dev. Difference	Rank P	Mean ChiSq	True ChiSq	Std. Dev. Difference	Rank P
15	1392.5	1619.8	3.189	3.00E-04	917.6	1088.4	4.76	1.00E-04	704.6	773	1.292	0.1
14	1264	1466.7	3.027	6.00E-04	856.8	1028.3	4.64	1.00E-04	646.8	711.5	1.299	0.1
13	1175.8	1403.5	3.309	2.00E-04	776.9	927.6	4.847	1.00E-04	595.8	651.9	1.169	0.13
12	1054.1	1250.6	3.15	6.00E-04	732.1	870.7	4.707	1.00E-04	577.6	632	1.142	0.14
11	894.6	1047.1	3.811	3.00E-04	679.7	828.9	4.968	1.00E-04	498.8	566.8	1.83	0.028
10	813.3	943.2	3.484	6.00E-04	507.7	616.8	2.281	8.40E-03	461.5	541.5	2.477	4.00E-03
9	738.5	876.4	3.671	1.00E-04	560.9	685.8	4.627	1.00E-04	421.4	480.8	1.633	0.46
8	641.9	764.7	3.535	2.00E-04	414.9	501.9	2.162	0.012	348.9	401.6	2.252	0.011
7	550.5	661.7	3.349	2.00E-04	427	518.2	4.037	1.00E-04	305.9	341.7	1.586	0.053

and reduced performance with 8 clusters. When using hairs only, the clusters were only significant when the samples were divided into 8, 10, or 11 clusters. This confirms the Random Forest findings that the swab samples provide stronger evidence for couple sharing than do the hair samples. With respect to the influence of the level of sexual activity, the couples who had sex an average of at least once during the seven days preceding collection tended to share a cluster

more often at that collection time point than the couples who had less sexual activity (Figure 3.5). While there was a general correlation between sexual activity and cluster sharing, there was no strong effect of increased frequency of sexual activity on increased clustering at a given time point above a certain activity level. For instance, both Couple 4 and Couple 12 averaged the same amount of sexual activity in the week prior to sample collection, but Couple 4 shared a cluster at

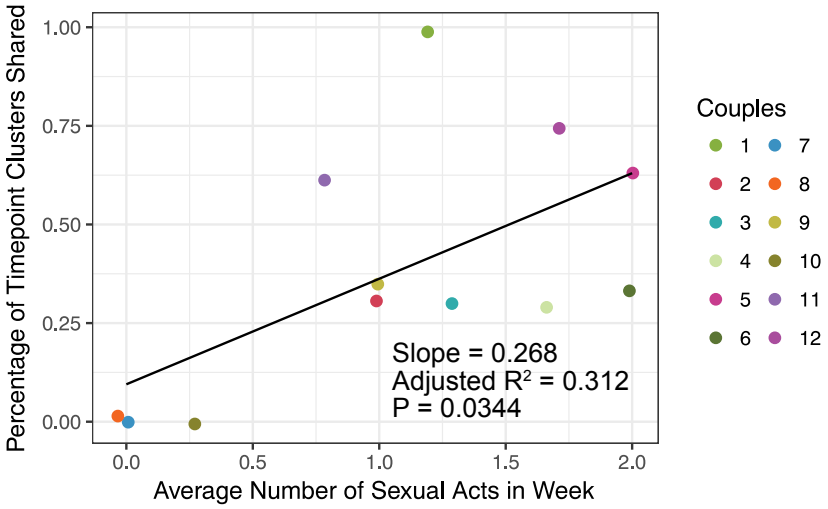


Figure 3.5 Comparison of average number of sexual encounters in week prior to collection and percentage of hair/swab clusters shared at each time point for each couple. The line of best fit using a linear regression model yields $R^2 = 0.31$, p=0.03.

33% of their time points while Couple 12 shared a cluster at 75% of their time points. This result establishes sexual activity per se is not sufficient to ensure microbiome sample sharing. With the elbow and ear samples, there was no correlation between sexual activity and clusters shared at a time point. A couple

with no sexual activity (couple 7) had the same amount of time point cluster sharing as a couple who averaged sexual activity twice a week. With the caveat that the sample size is small, the clustering of couples using the ear and elbow samples does not appear to be correlated with sexual activity, so may be due to some other mechanism.

3.4.6 Closest Non-self Neighbor

While I was analyzing my dataset, Ross et al. [44] reported a similar analysis of the effect of cohabitation on skin microbiome sharing across 17 sites in 10 sexually active couples. They assessed which non-self sample was nearest to each person in the beta diversity distance matrix space, asking how often it was their partner or an unrelated individual. Applying this approach to my dataset, an individual's sample was closest to another of their own samples the majority of the time (Figure 3.6 a). Using the weighted UniFrac distances, an individual's own sample was the closest sample for 74.9% of pubic area samples, 92.1% of elbow samples, and 69.7% of ear samples. The highest within-individual closest matches for Ross et al. were for the thigh (89.7%) and eyelid (77.5%). When compared solely to non-self samples, all three skin areas had similar proportions of closest match samples belonging to a partner (Figure 3.6 b) with the exception of the weighted UniFrac elbow samples, which had no closest partner samples. The percentage of samples with partner closest-match samples (0-9.2% weighted, 7.9-11.6% unweighted) was lower than Ross et al. found with the torso, navel, and eyelid (17.5%-21%). To account for the effect from individuals who did

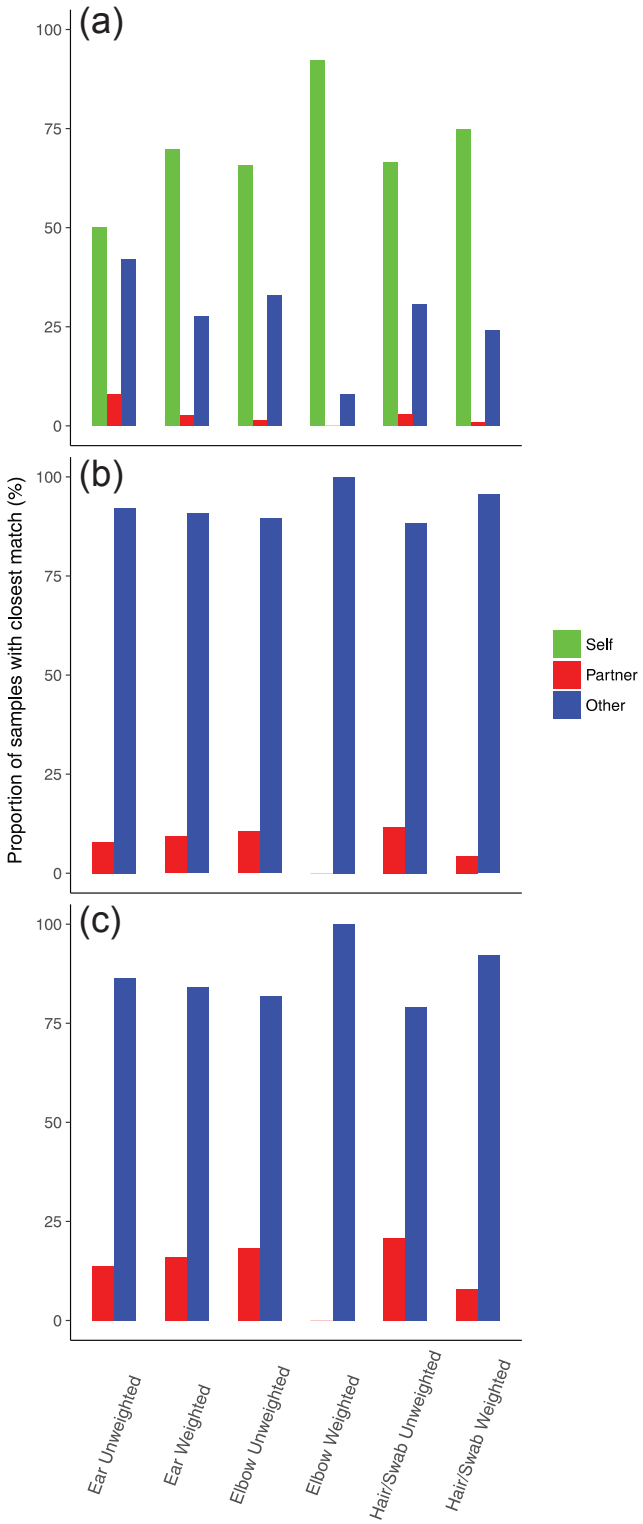


Figure 3.6 Proportion of closest-neighbor samples using Weighted and Unweighted UniFrac distances for ear, elbow, and pubic hair/swab samples for (a) self, partner, and others, (b) just partner or other, (c) and just partner or other for just the couples.

not have a partner in my data, closest non-self neighbor analysis was repeated for just the couple samples (Figure 3.6 c). The pattern remained the same while the proportion of partner closest-neighbor matches increased. Generally, the individuals were closest to their partner with pubic area samples when analyzed using the unweighted UniFrac distance. The individuals who had samples with their partner as nearest neighbor using weighted UniFrac were not the same individuals with partner nearest neighbor samples using unweighted UniFrac (Supplemental Tables 3.2-3.4). Within a couple, the partners did not always have reciprocal sharing and varied in the number of times the closest neighbor of an individual was their partner. The couples who did not have sex in the seven days prior to collection for all of their pubic area time points (couples 7 and 8) did not have their partner as their closest neighbor for any of their samples. There also were sexually active couples who not did have their partner as their closest neighbor for any of their samples (Supplemental Table 3.2). The results here are in agreement with Ross et al.'s finding that Random Forest modeling performs better than closest matching sample method for predicting couples.

3.5 Discussion

The results presented here demonstrate the potential for using the microbiome from the pubic mound area, whether collected from hairs or with swabs, as a tool for comparing individuals, with implications for use in forensic investigations. As

has been demonstrated in other body areas [15, 16, 45, 46], the pubic mound area microbiome shows the ability to differentiate one individual from another individual. When looking at the effect sizes of various sample categories on the beta diversity, “Individual” was the largest driver of variation in the samples. Use of Random Forest classification models proved to be powerful in predicting the “Individual” labels of samples. When a subset of hairs was re-sampled and re-processed, this model correctly predicted the “Individual” label in 9 out of 10 instances. The same model uniquely labeled two of the prior storage study participants with individuals known to have participated in that study and the current study. All of these results extend my previous demonstration that technical replicates of pubic microbiome samples provide consistent individualization by showing that biological replicates obtained over intervals of weeks and months are also somewhat individualized.

Even with the potential of mixtures of microbiomes with sexually active subjects, the estimated error rate in “Individual” prediction was low. The predicted error rate in the Random Forest classification model was highest when hairs alone was used. The best predicted error rate was found when just swabs were used with the predicted error rate of using hairs and swabs together falling between hairs and swabs. Given the similar numbers of swabs and hairs, this change in error rate is more likely to be due to higher variation of the microbiome in the single hairs of an individual compared to the swabs as opposed to more samples evening out variation while constructing the classification model. While the

distribution of read counts between the swabs and hairs was significantly different (*t*-test, $p = 1.85 \times 10^{-5}$), the means and medians of each (hairs: 107,494/73,367 reads and swabs: 300,160/165,699 reads, respectively) show that there should be sufficient reads from the hairs such that the alpha diversity of the hairs should be represented as well as with the swabs. As the alpha diversity differences between different lengths of hairs used for extractions was not significant, significantly more hair, length or quantity, may need to be used in order to increase the detected alpha diversity and even out differences between sampling efforts. Even so, a mixed pool of pubic hair and pubic swab microbiomes or just pubic swab microbiomes presents the potential for identification of an individual. In the next chapter I explore the forensic implications of this finding.

Couples also tended to be more similar to each other than to non-related individuals. When looking at the beta diversity effect sizes, the “Couple” category had the second-largest effect size for both the weighted and unweighted UniFrac measures. Because the unweighted UniFrac distances are more sensitive to differences in the lower abundance OTUs, it was thought that the unweighted UniFrac would reveal if couples were more alike in these OTUs. The results did not support this supposition as groupings by couple using the unweighted UniFrac distances performed no better. As these effect sizes were determined in a univariate manner and non-coupled participants were included as “couples” in the analysis and where couple = individual for these participants, the effects of the “Individual” may also be playing a role in the “Couple” effect size. That is, the

twelve couple pairs were couples 1-12 while the remaining individuals were couples 13-31 such that there were more individuals contributing to the “Couple” category than actual couples. However, the additional analyses by hierarchical clustering also strongly support the similarities between the couples.

The similarity of couples need not be due to physical contact, since other studies have shown that co-habiting individuals tend to share microbiome profiles at a variety of skin sites. I thus asked whether the couple sharing is greater for the pubic region than other sites, and found somewhat surprisingly that this is not clearly the case. The ability to predict a couple using the Random Forest classification model was similar for both the intimate (pubic region) and less-intimate (inner elbow and behind the ear) samples. However, there may be different mechanisms in place that account for the similarities within a region. It may be that the shared environment and contact between the couples carries over to the less intimate areas of the body either through casual touching, secondary transfer or some other unknown mechanism, creating related microbiome profiles. In one study, this similarity extended to the family dog [18]. Another study has shown relatedness between couples at multiple skin sites, including the eyelids [44], an area much like behind the ear where direct contact between corresponding body sites is not expected. They also observed greater similarity between areas like the torso and inner thigh which are more likely to come into contact. It is unknown how much time in a shared environment or how much contact is required before convergence of microbiomes is detectable. This,

as well as the longevity of profiles after co-habitation ceases or following a single encounter, is worthy of further study.

The hierarchical clustering results also showed a general trend for couples to be more similar in their pubic hair/swabs microbiomes. There was no direct correlation between when a couple last had sex, how long after sex a sample was collected, and when the individuals in the couple shared a cluster, but there was a minimum threshold of sexual activity before couples tended to be more similar to one another through shared clustering of pubic area samples. The elbow and ear samples provide preliminary evidence that a different transfer mechanism as cluster sharing was not dependent on sexual activity, with the non-active couples sharing the same number of clusters as the most-active couples. The incomplete nature of the microbiome sharing even for highly sexually active couples suggests that the detection of transfer during a single sexual encounter, such as rape, is unlikely. It may occur in some cases, but clearly cannot be expected to provide positive identification in most cases.

Closest neighbor analyses showed inconsistent results from couple-to-couple with a non-partner sample being the closest non-self sample the majority of the time. While the general trend is that only couples who were sexually active had their partner as their nearest neighbor for some of their pubic area samples was replicated, not all of the sexually active couples exhibited this pattern. Unlike the previous analyses, the unweighted UniFrac distances, where rarer bacteria exhibit

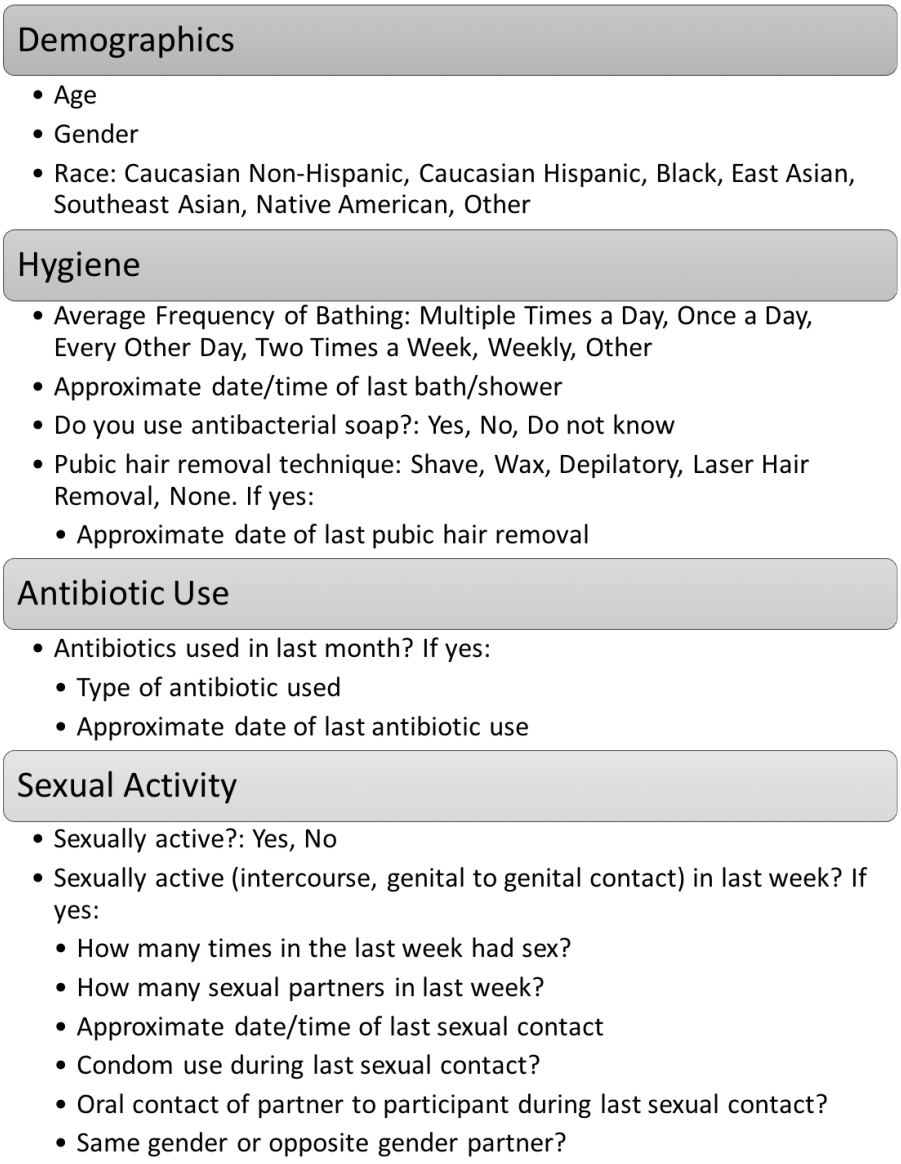
larger effects due to relative abundances not being included, exhibited more partners as closest neighbors. When including relative abundances, as with the weighted UniFrac, subtle sharing of rarer bacteria was lost. As with the hierarchical clustering, the couples who were not sexually active in the week prior to collection for all of their samples did not have their partner as their nearest neighbor for any of their samples. Even though we did not have oral samples from the participants, the samples where oral contact was indicated were not any closer to a publically available buccal mucosal data set, which was distinct from the pubic hair/swab samples, than samples where no oral contact was indicated. Given the rapidly diminishing ability to detect the transfer of an inoculation of bacteria in prior work [17], the difficulty in detecting a transfer event from oral to pubic area with much lower amounts of material is unsurprising. The persistence of tongue microbiome on skin surfaces has also been shown to be dependent on the type of skin surface, with the inoculated microbiome exhibiting longer persistence on sebaceous skin sites (the forehead) as compared to dry skin sites (forearm) [15]. Here, the pubic hair/swabs would not be expected to have as much turnover as the mouth where saliva is constantly rinsing the area and the pubic hair/swabs were more similar to sebaceous skin than dry skin but I was unable to detect the transfer. There may have been more direct contact between the mouth and the genital region so that not as much oral microbiome was transferred to the upper pubic mound region. It is also likely that washing between sex and sampling for many of the cases will have diluted any transfer, whereas sampling soon after the event and without washing might be more informative.

While this study was observational and the participants were not instructed to behave in any particular manner, I was able to show the great potential of individualization using the pubic hair/swab microbiome, also with some propensity to group couples. Controlled studies involving collection of samples immediately prior to sexual contact and then at fixed time points after sexual contact would serve to limit some of the variability inherent in this study and clarify the ability to detect transfer events. These studies would though be logistically difficult, as would studies designed to longitudinally follow the consequences of a single sexual encounter with a novel partner.

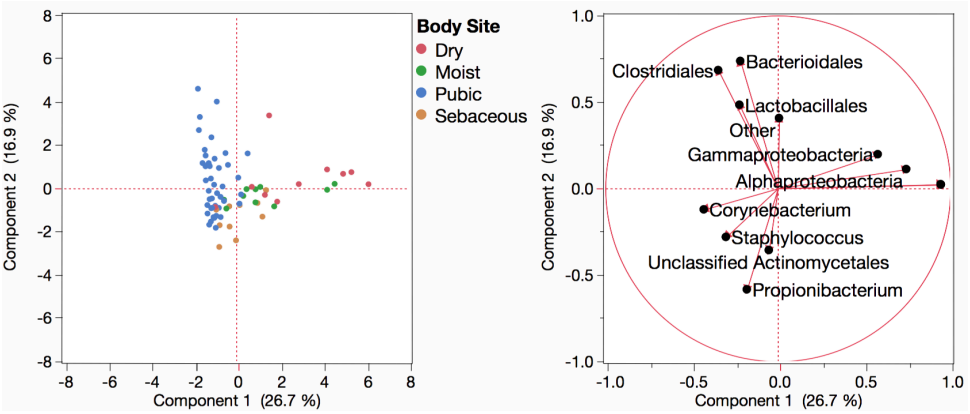
Given criticism of microscope hair comparison techniques such as lack of foundational validity and lack of consistent reporting of the weight of hair associations [6], development of techniques for the analysis of hair in forensic investigations that have undergone rigorous scientific testing would allow for the potential to use an underutilized piece of evidence. I have taken the first steps toward the development of such a technique, using the microbiome of the pubic hair. While characterization of the pubic hair/pubic area microbiome using the V3/V4 region of the 16S rRNA gene does not appear to be sensitive enough to detect individual sexual events, it may be possible to detect more differences or similarities using something like whole genome amplification and metagenomics. Obstacles such as working with low biomass samples will need to be overcome in order to use techniques typically used with higher quantities of sample. Currently,

this technique shows promise in matching an individual by using the microbiome. Further work will confirm these results with a larger sample set and refine the necessary statistics needed to provide weight to the evidence. As a proof-of-concept, the data here are the first step towards a new forensic technique.

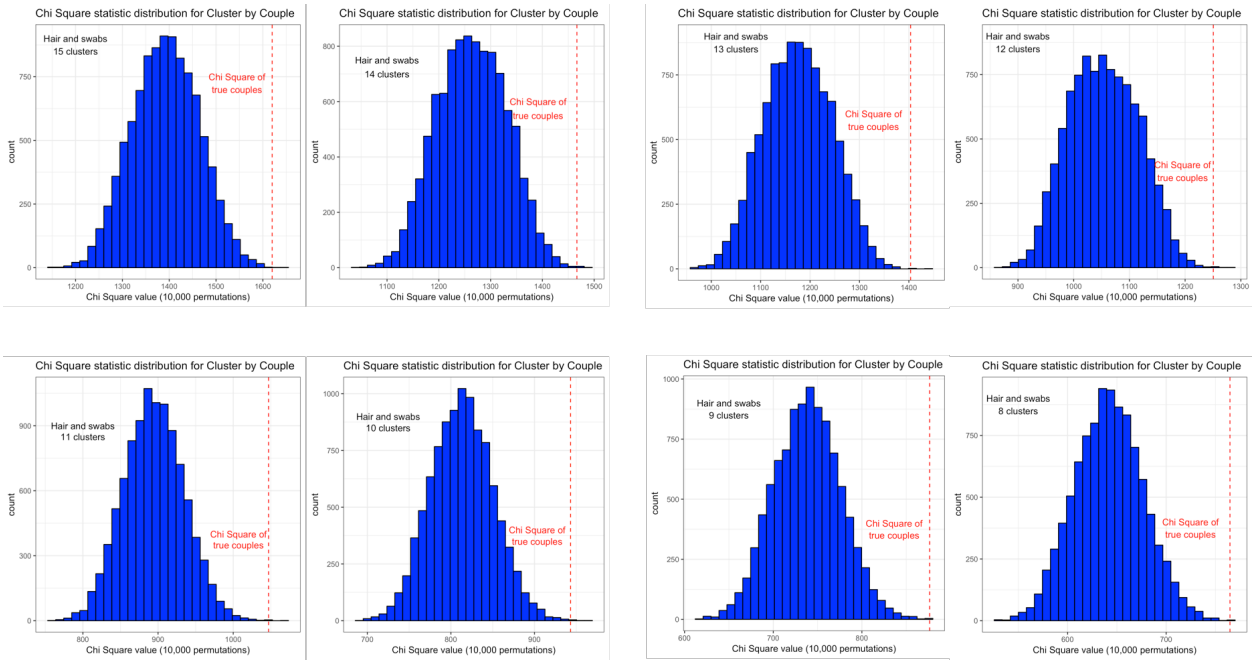
3.6 Supplemental Figures and Tables



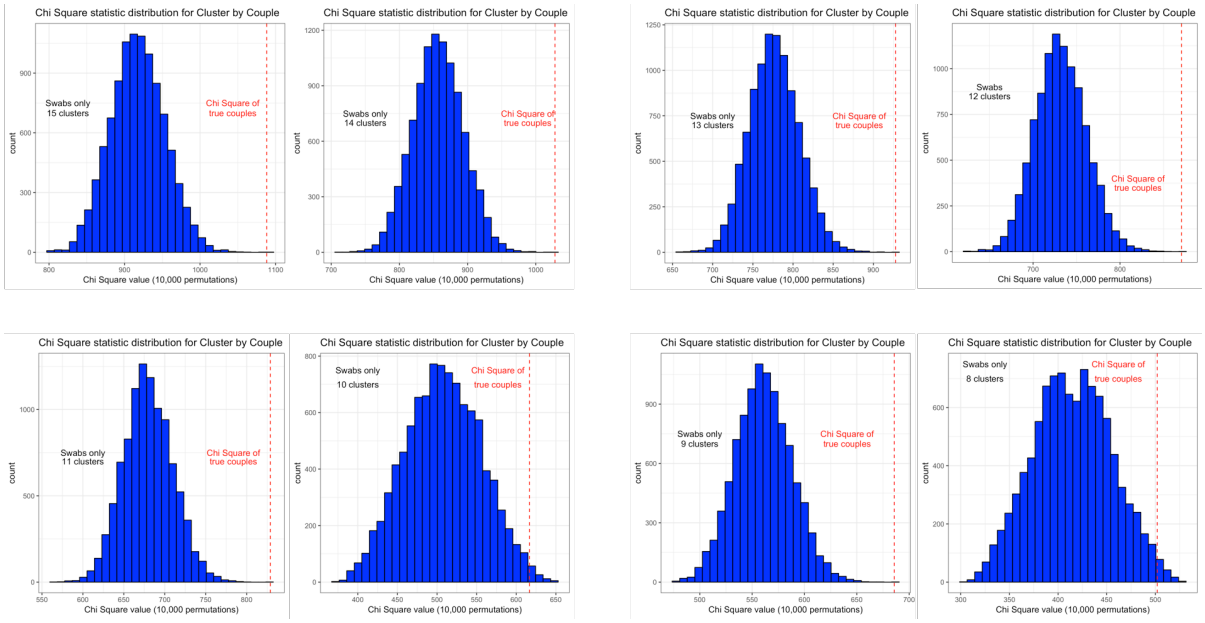
Supplemental Figure 3.1 Contents of survey questions completed by participants at each collection time point



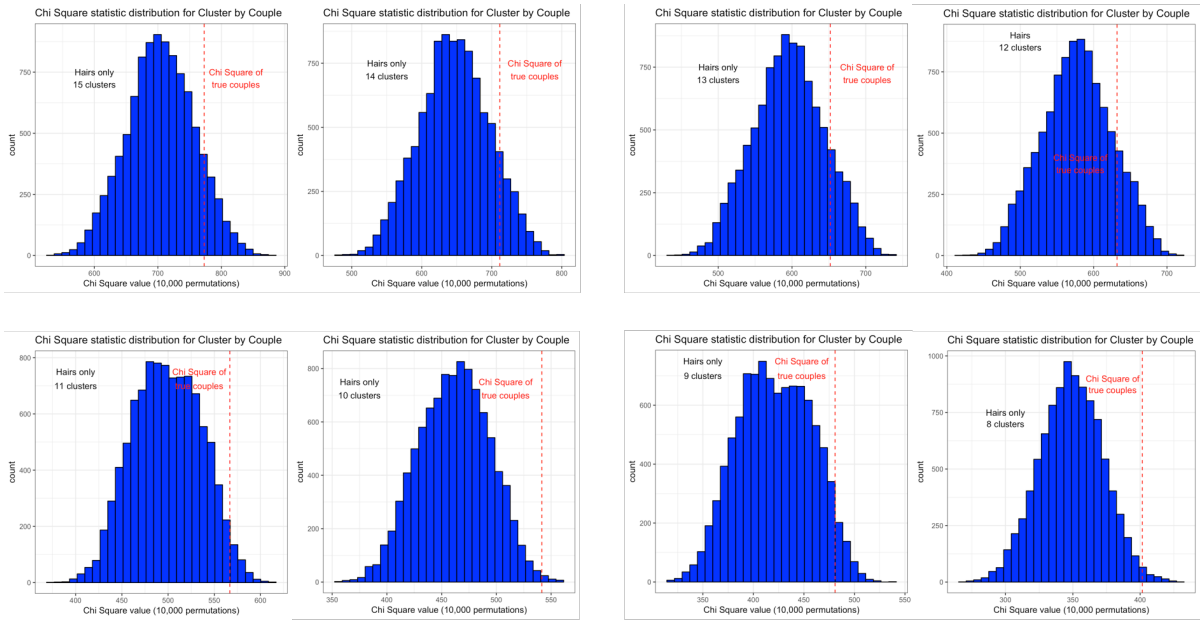
Supplemental Figure 3.2 Principal components analysis showing the relationship among microbiomes of various skin sites from Grice et al. [42] based on taxon relative abundances. The genera that are driving the differences are represented in the plot on the right.



Supplemental Figure 3.3 Chi-square distributions of hair and swabs clusters for 8-15 clusters. The dotted line represents the Chi Square value for the true couples.



Supplemental Figure 3.4 Chi-square distributions of swabs-only clusters for 8-15 clusters. The dotted line represents the Chi Square value for the true couples.



Supplemental Figure 3.5 Chi-square distributions of hair-only clusters for 8-15 clusters. The dotted line represents the Chi Square value for the true couples.

Supplemental Table 3.1 Metadata of HMP buccal mucosal samples

Sample	Gender	Read Count	File Link (http://downloads.hmpdacc.org/data/HM16STR/)
Oral1	male	14979	by_sample/SRS015712.fsa.gz
Oral2	female	4642	by_sample/SRS016461.fsa.gz
Oral3	male	10145	by_sample/SRS016865.fsa.gz
Oral4	male	6484	by_sample/SRS017410.fsa.gz
Oral5	female	5713	by_sample/SRS017780.fsa.gz
Oral6	male	6948	by_sample/SRS018092.fsa.gz
Oral7	female	5695	by_sample/SRS018208.fsa.gz
Oral8	female	4300	by_sample/SRS022798.fsa.gz
Oral9	female	1	by_sample/SRS022834.fsa.gz
Oral10	female	3800	by_sample/SRS022876.fsa.gz
Oral11	male	2	by_sample/SRS023334.fsa.gz
Oral12	male	5713	by_sample/SRS023365.fsa.gz
Oral13	female	5216	by_sample/SRS023389.fsa.gz
Oral14	female	6776	by_sample/SRS023689.fsa.gz
Oral15	female	6768	by_sample/SRS023809.fsa.gz
Oral16	male	9412	by_sample/SRS024527.fsa.gz
Oral17	female	11066	by_sample/SRS045353.fsa.gz
Oral18	male	10784	by_sample/SRS056841.fsa.gz
Oral19	male	18949	by_sample/SRS058695.fsa.gz
Oral20	female	10460	by_sample/SRS064652.fsa.gz
Oral21	male	17530	by_sample/SRS064715.fsa.gz
Oral22	male	7219	by_sample/SRS065316.fsa.gz
Oral23	male	8787	SRP002860/SRS064807.fsa
Oral24	male	15909	SRP002860/SRS077304.fsa
Oral25	female	6172	SRP002860/SRS078514.fsa
Oral26	male	19336	SRP002860/SRS097899.fsa
Oral27	female	12660	SRP002860/SRS100781.fsa
Oral28	male	2709	SRP002860/SRS104283.fsa
Oral29	male	1468	SRP002860/SRS104337.fsa
Oral30	female	2923	SRP002860/SRS104682.fsa
Oral31	female	2422	SRP002860/SRS105054.fsa
Oral32	female	3439	SRP002860/SRS143476.fsa
Oral33	male	6317	SRP002860/SRS144162.fsa
Oral34	male	4626	SRP002860/SRS144420.fsa
Oral35	male	3470	SRP002860/SRS145343.fsa
Oral36	male	5533	SRP002860/SRS146790.fsa
Oral37	female	7384	SRP002860/SRS147145.fsa
Oral38	female	5106	SRP002860/SRS147356.fsa
Oral39	female	8504	SRP002860/SRS148118.fsa
Oral40	female	4592	SRP002860/SRS148997.fsa

Supplemental Table 3.2 Pubic area summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances

Couple	Participant Code	Pubic Swabs/Hairs Weighted				Pubic Swabs/Hairs Unweighted			
		Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage	Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage
1	PF10	10	0			10	4	40.0	40.0
	PM45	10	0			10	4	40.0	
2	PM72	10	0			10	0		
	PF56	10	0			10	0		
3	PM04	11	0			11	0		
	PF21	9	0			9	0		
4	PM14	10	0			12	0		16.7
	PF46	12	0			12	4	33.3	
5	PM31	20	8	40.0	36.7	20	11	55.0	53.3
	PF19	10	3	30.0		10	5	50.0	
6	PF37	10	0			10	1	10.0	4.8
	PM83	11	0			11	0		
7	PM61	12	0			12	0		
	PF11	12	0			12	0		
8	PF74	6	0			6	0		
	PM77	8	0			8	0		
9	PM73	20	1	5.0	11.1	20	1	5.0	8.3
	PF28	16	3	18.8		16	2	12.5	
10	PF92	8	0			8	2	25.0	21.4
	PM53	6	0			6	1	16.7	
11	PF67	20	0			20	10	50.0	35.9
	PM24	19	0			19	4	21.1	
12	PM80	16	3	18.8	21.9	16	4	25.0	37.5
	PF26	16	4	25.0		16	8	50.0	
Total		292	22	7.5		294	61	20.7	

Supplemental Table 3.3 Elbow summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances

		Elbow Weighted				Elbow Unweighted			
Couple	Participant Code	Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage	Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage
1	PF10	4	0			4	1	25.0	12.5
	PM45	4	0			4	0		
2	PM72	4	0			4	1	25.0	100.0
	PF56	4	0			4	0		
3	PM04								
	PF21								
4	PM14								
	PF46								
5	PM31								
	PF19								
6	PF37	4	0			4	4	100.0	62.5
	PM83	4	0			4	1	25.0	
7	PM61	4	0			4	0		
	PF11	4	0			4	0		
8	PF74								
	PM77								
9	PM73								
	PF28								
10	PF92								
	PM53								
11	PF67								
	PM24								
12	PM80	4	0			4	0		
	PF26	4	0			4	0		
Total		40	0			40	7	17.5	

Supplemental Table 3.4 Ear summary of closest non-self neighbor samples for couples using weighted and unweighted UniFrac distances

		Ear Weighted				Ear Unweighted			
Couple	Participant Code	Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage	Number of Samples	Partner Matches	Match Percentage	Couple Match Percentage
1	PF10	4	0			4	3	75.0	62.5
	PM45	4	0			4	2	50.0	
2	PM72	4	0			4	1	25.0	12.5
	PF56	4	0			4	0		
3	PM04								
	PF21								
4	PM14								
	PF46								
5	PM31								
	PF19								
6	PF37	4	0		37.5	4	0		
	PM83	4	3	75.0		4	0		
7	PM61	4	2	50.0	37.5	4	0		
	PF11	4	1	25.0		4	0		
8	PF74								
	PM77								
9	PM73								
	PF28								
10	PF92								
	PM53								
11	PF67								
	PM24								
12	PM80	4	0			4	0		
	PF26	4	0			4	0		
Total		40	6	15.0		40	6	15.0	

3.7 References

[1] World Health Organization. Global and regional estimates of violence against women: prevalence and health effects of intimate partner violence and non-partner sexual violence. 2013.

[2] Truman JL, Langton L. Criminal Victimization, 2014. In: U.S. Department of Justice OoJP, Bureau of Justice Statistics, editor.2015.

[3] Department of Defense. Department of Defense Annual Report on Sexual Assault in the Military. Fiscal Year 2015.

[4] National Crime Victimization Survey, 2010-2014. In: Department of Justice OoJP, Bureau of Justice Statistics, editor.2015.

[5] National Incident-Based Reporting System, 2012-2014. In: Investigation FBo, editor.2015.

[6] President's Council of Advisors on Science and Technology. Report to the President: Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-Comparison Methods. 2016.

[7] Blum HE. The human microbiome. Advances in Medical Sciences. 2017;62:414-20.

[8] Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, et al. Structure and function of the global ocean microbiome. Science. 2015;348:1261359.

[9] DeLeon-Rodriguez N, Lathem TL, Rodriguez-R LM, Barazesh JM, Anderson BE, Beyersorf AJ, et al. Microbiome of the upper troposphere: Species composition and prevalence, effects of tropical storms, and atmospheric implications. PNAS. 2013;110:2575-80.

[10] Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project: successes and aspirations. BMC Biology. 2014;12:69.

[11] Sha S, Liang J, Chen M, Xu B, Liang C, Wei N, et al. Systematic review: faecal microbiota transplantation therapy for digestive and nondigestive disorders in adults and children. Alimentary Pharmacology and Therapeutics. 2014;39:1003-32.

[12] Mimee M, Citorik RJ, Lu TK. Microbiome therapeutics - Advances and challenges. Advanced Drug Delivery Reviews. 2016;105:44-54.

[13] The NIH HMP Working Group. The NIH Human Microbiome Project. Genome Research. 2009;19:2317-23.

[14] The Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. Nature. 2012;486:207-14.

[15] Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. Science. 2009;326:1694-7.

[16] Califf K, Gonzalez A, Knight R, Caporaso JG. The Human Microbiome: Getting Personal. Microbe. 2014;9:410-5.

[17] Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. Microbiome. 2014;2:41.

[18] Song SJ, Lauber CL, Costello EK, Lozupone CA, Humphrey GC, Berg-Lyons D, et al. Cohabitating family members share microbiota with one another and with their dogs. eLife. 2013;2:e00458.

[19] Metcalf JL, Xu Z, Bouslimani A, Dorrestein P, Carter DO, Knight R. Microbiome tools for forensic science. Trends in Biotechnology. 2017.

[20] Clarke TH, Gomez A, Singh H, Nelson KE, Brinkac LM. Integrating the microbiome as a resource in the forensics toolkit. Forensic Science International: Genetics. 2017;30:141-7.

[21] Metcalf JL, Parfrey LW, Gonzalez A, Lauber CL, Knights D, Ackermann G, et al. A microbial clock provides an accurate estimate of the postmortem interval in a mouse model system. eLife. 2013;2:e01104.

[22] Hyde ER, Haarmann DP, Lynne AM, Bucheli SR, Petrosino JF. The living dead: bacterial community structure of a cadaver at the onset and end of the bloat stage of decomposition. PLoS ONE. 2013;8:e77733.

[23] Hyde ER, Haarmann DP, Petrosino JF, Lynne AM, Bucheli SR. Initial insights into bacterial succession during human decomposition. Int J Legal Med. 2015;129:661-71.

[24] Moreno LI, Mills D, Fetscher J, John-Williams K, Meadows-Jantz L, McCord B. The application of amplicon length heterogeneity PCR (LH-PCR) for monitoring the dynamics of soil microbial communities associated with cadaver decomposition. Journal of Microbiological Methods. 2011;84:388-93.

[25] Carter DO, Metcalf JL, Bibat A, Knight R. Seasonal variation of postmortem microbial communities. Forensic Sci Med Pathol. 2015;11:202-7.

[26] Yao X, Liu W, Han J, Pei G, Tong Y, Luo Y. Analysis of microbiome DNA on frequently touched items and from palm prints. J Forensic Sci Med. 2016;2:74-7.

[27] Wilkins D, Leung MH, Lee PK. Microbiota fingerprints lose individually identifying features over time. Microbiome. 2017;5.

[28] Tridico S, Murray DC, Addison J, Kirkbride KP, Bunce M. Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investigative Genetics. 2014;5.

[29] Williams DW, Gibson G. Individualization of pubic hair bacterial communities and the effects of storage time and temperature. Forensic Science International: Genetics. 2017;26:12-20.

[30] Illumina. 16S Metagenomic Sequencing Library Preparation. Part # 15044223 Rev. B ed.

[31] http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

[32] Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. PANDAseq: paired-end assembler for Illumina sequences. BMC Bioinformatics. 2012;13.

[33] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7:334-6.

[34] Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics. 2011;27:2194-200.

[35] Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics. 2010;26:266-7.

[36] DeSantis T, Hugenholtz P, Larsen N, Rojas M, Brodie E, Keller K, et al. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl Environ Microbiol. 2006;72:5069-72.

[37] Anderson MJ. A new method for non-parametric multivariate analysis of variance. Austral Ecology. 2001;26:32-46.

[38] Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. eSystems. 2017;2:e00191-16.

[39] Breiman L. Random Forests. Machine Learning. 2001;45:5-32.

[40] Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2:18-22.

[41] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

[42] Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC, et al. Topographical and Temporal Diversity of the Human Skin Microbiome. Science. 2009;324:1190-2.

[43] The Human Microbiome Project Consortium. A framework for human microbiome research. Nature. 2012;486:215-21.

[44] Ross AA, Doxey AC, Neufeld JD. The skin microbiome of cohabiting couples. mSystems. 2017;2:e00043-17.

[45] Fierer N, Lauber CL, Zhou N, McDonald D, Costello EK, Knight R. Forensic identification using skin bacterial communities. PNAS. 2010;107:6477-81.

[46] Goga H. Comparison of bacterial DNA profiles of footwear insoles and soles of feet for the forensic discrimination of footwear owners. Int J Legal Med. 2012;126:815-23.

CHAPTER 4

MIXTURES

4.1 Abstract

Many forensically relevant biological samples consist of mixtures of material from two or more individuals. The ability to detect and interpret these mixtures in an investigation of sexual assault is important for the inclusion of a perpetrator in support of a prosecution, support of the victim in resolving what occurred during an incident, and the exoneration of individuals as participants in an incident. As I have shown in Chapter 3, the pubic hair/pubic area microbiome performs well in identifying individuals with similarities between couples. For the scenario of detecting a mixture of individuals as a result of a single incident of sexual contact, it is still unclear under what conditions the pubic hair/pubic area microbiome can be used to show that there is a mixture present and how to include or exclude an individual from that mixture. In order to explore various methods of mixture interpretation, *in silico* couples were created from the population data set from Chapter 3 and analyzed using percent OTU sharing, UniFrac distances, Random Forest modeling, and SourceTracker analysis. Analyses here imply a minimum of 5% contribution from a donor to be able to detect any change in the host's microbiome with at least 25% contribution for reliable detection. Using Deblur-assigned OTUs, high sensitivity and specificity was achieved for detection of a

donor individual above a contribution level above 25% with at least 55% OTU sharing. These simulations constitute the first step towards the development of a mixture analysis scheme, allowing for the utilization of pubic hairs/pubic area swabs in sexual assault investigations where other avenues of analysis, such as traditional DNA typing of semen, are unavailable.

4.2 Introduction

Identification using the pubic hair/pubic microbiome has forensic applications, such as determining the identity of an individual through a pubic hair left behind at a scene. This ability is useful in cases where the root of the hair is not available for traditional forensic short tandem repeat (STR) testing or as a compliment to human mitochondrial DNA (mtDNA) testing. However, the pubic hair/pubic area microbiome would be truly powerful if it were able to detect recent sexual activity through the exchange of the microbiome. Tridico et al. suggested that it is possible to detect when sexual contact has occurred through the pubic hair microbiome [1]. They were able to see a shift in beta diversity in two individuals they later determined had sexual contact prior to the sample collection. Additional work shows that sexually active couples tend to be more similar to each other than unrelated individuals but the ability to consistently detect single instances of sexual contact was not supported (Chapter 3).

The task of detecting and interpreting mixtures of microbiomes is not trivial. By contrast, determining whether there is a mixture of human STRs is a relatively simple matter of detecting more than two alleles at multiple loci. Additionally, an individual's STR profile does not change over time. Interpretation of a STR mixture may be more difficult depending on the number and proportion of contributors in the mixture, but the fact that there is a mixture is usually not difficult to determine. The microbiome of a single individual is already a mixture of bacteria with normal fluctuations over time in both species composition and relative abundances of those species. Additionally, at a given body site, many of the bacterial species are shared between individuals. It has been shown by Kort et al. that it is possible to detect the transfer of an inoculation of marker bacteria through kissing [2]. In this case, a large number of bacteria not part of the predominant oral genera was spiked into the donor community via a yogurt-based beverage and was used for detecting the transfer. There was a sharp decline in the quantity of marker bacteria detected after each transfer. When considering bacteria that are part of the normal flora, how much of the change in an individual's pubic area microbiome are due to normal fluctuations and how much these changes are due to exchange of the microbiome between individuals during contact or some other perturbation remains to be seen. In order to detect exchange of bacteria between individuals, it must first be determined what proportion of the donor microbiome needs to be exchanged in order to have the statistical power to detect this transfer.

During mixture interpretation, additional consideration must be paid to the question being asked. While it is not possible to determine whether a sample originates from a mixture of individuals on the basis of that sample alone, it may be possible to use a Bayesian framework to determine the probability that a sample represents a mixture derived from one or more proposed individuals, for example suspects in a sexual assault case. Potential scenarios include the following: what is the likelihood of a given microbiome profile if a suspect is a contributor? What is the likelihood of a given microbiome profile if any or all individuals from a pool of suspects are contributors? What is the power to exclude one individual in a pool of suspects from likely contribution to the mixture? How well these models work will depend, in part, on the sensitivity and precision of the test.

To this end, this Chapter describes *in silico* simulations designed to evaluate the power to detect mixtures of pubic hair/area microbiomes. Mixtures of individuals' microbiomes were created in varying ratios, allowing me to control how much each person is contributing to the mixtures. Shifts in beta diversity as well as how these mixtures fared with Random Forest modeling were explored. The results place lower bounds on mixture detection in real-world scenarios where *in vivo* mixtures would have unknown contributions from each individual even when we know contact has occurred.

These *in silico* predicted relationships between individuals in mixtures were then compared with the percentages of operational taxonomic units (OTUs) shared in pairwise fashion between types of samples from Chapter 3. As a parallel mode of analysis, I also computed the pairwise weighted UniFrac distances between samples. Various categorical groupings were used, such as pseudo-Couples versus Individuals. These same analyses were extended to the larger population samples to incorporate the effects of variation in the microbiome in an individual over time.

Preliminary Bayesian analyses were then conducted to evaluate the ability to predict whether an individual is present in a sample. These analyses include the sensitivity and specificity under various conditions and scenarios of prior mixture probabilities, which influence the posterior probabilities much as precision is a function of the true proportion of the mixtures. I also evaluated the ability to identify specific contributors to a mixture using SourceTracker, a tool utilizing a Bayesian approach to identify the sources and proportions of contamination [3]. In the case of forensic-type samples, like the pubic hair/pubic area swabs here, the suspect or a pool of suspects could be considered a potential source of "contamination" in the sample. This tool also considers an unknown source as potentially contributing to a sample such that the source pool is not limited to those individuals of whom we are aware. These techniques require a known reference sample from each individual used for comparison and would need to be

a sample where it can be reasonably assumed that the sample is not a mixture of individuals.

These techniques will show the growing potential and uses of the human pubic hair/pubic area microbiome as a forensic tool in cases of sexual contact. Through some of these analyses, additional areas of inquiry are revealed that would serve to strengthen the analyses of microbiome mixtures. These inquiries will inform the analyses of any microbiome mixture, not just the pubic hair/pubic area. A deeper discussion of additional requirements before any diagnostic can be incorporated into forensic case work, as well as of new avenues of research, follows in the final chapter.

4.3 Materials and Methods

Artificial mixtures were created *in silico* using custom scripts to sample at random from the QIIME 1 (QI) OTU data collected in Chapter 3 in order to test the ability to detect mixtures of individuals using various computational methods. Opposite-sex "pseudo-couples" were created using all 20 female participants and 20 of the male participants, being careful not to pair known couples. The microbiome profile from the first time point was used for each individual by collapsing the OTU table by "first" and relative abundances of the OTUs using QIIME1 [4] (*collapse_samples.py --collapse-mode first --normalize*). Relative abundances

were used at this step to reduce variation from differences in read depths. The relative abundances of each OTU were then multiplied by 10,000 to give a read depth of 10,000 reads for each sample. The following mixture ratios were created for each couple (A:B): 100:0, 99:1, 95:5, 90:10, 75:25, 66:33, 50:50, 33:66, 25:75, 10:90, 5:95, 1:99, and 0:100. In order to model stochastic sampling as would be expected *in vivo* where lower abundance OTUs would be sampled at a more variable rate, individuals were rarefied without replacement (*single_rarefaction*.py) at the following levels to correspond to ratio contributions to a 10,000 read mixture: 9900, 9500, 9000, 7500, 6660, 5000, 3340, 2500, 1000, 500, and 100 reads. These rarefied samples were then combined for each couple at the above ratios.

Sequences that were processed using Deblur [5] in order to better resolve sequence differences at a finer resolution were also used to create *in silico* mixtures to explore whether this finer resolution would further define mixtures of couples from unrelated individuals. Here, the OTU table was collapsed by individual and summing the OTUs then rarefying the resulting sequences to 10,000 reads. These reads were used to make the mixture ratios as above.

Shifts in beta diversity were determined using the weighted and unweighted UniFrac [6, 7] through the *beta_diversity.py* script in Q1 on each “couple” mixture and on these “couples” in a combined set. The beta diversity was used to

calculate the principal coordinates (*principal_coordinates.py*) and make PCoA plots (*make_2d_plots.py*).

The Random Forest model created with the original population collection in Chapter 3 was used to predict the 'Individual' designation in the Q1 *in silico* couples using the randomForest R package [8]. Because it has been previously observed with the original samples that Deblur did not significantly improve the performance of the beta diversity and Random Forest modeling, the *in silico* Deblur mixtures were only used here for the percent OTU sharing to compare its performance with the QI OTUs. The creation of this model involved the construction of a specified number of classification trees. Here, the default number of 500 trees was used. A matrix of probabilities was created by passing type="prob" in the *predict* command. These probabilities represented the proportion of the time a given 'Individual' designation was predicted in the 500 classification trees created in setting up the original model. For each couple, it was determined which individuals were predicted first and second most often and what rank each of the individuals in the mixture were when ordering the probabilities from largest to smallest. I also determined from the Random Forest model which OTUs were ranked highest in importance for the model for each Individual and the average relative abundance of that OTU for that Individual.

Based upon the results of the predictions, it became apparent that the results were biased by the presence of "dominant" individuals whose microbiome profile

tended to make an inflated contribution, often attributable to one or a few highly abundant taxa. Two strategies were developed to deal with this bias. First, the dominant individuals were determined by ascertaining which individuals consistently incorrectly scored as first or second match in the mixtures. These individuals were removed from the original OTU table and a new Random Forest model was created with the randomForest R package. The *in silico* couple mixtures containing those Individuals were removed and a new probability matrix was created for the remaining mixtures with the *predict* command using this new Random Forest model. Second, the top high importance OTU for each Individual was filtered from the original OTU table and a new Random Forest model was created from this new OTU table. This adjusted Random Forest model was used with the entire *in silico* mixture set to make an updated probability matrix. As before, it was determined which individuals were predicted first and second most often in each couple set and what rank each of the individuals in the mixture were when ordering the probabilities from largest to smallest.

In order to explore the relatedness of samples as a function of OTUs shared, the number of unique OTUs was determined for each single individual in each *in silico* data set. The number of OTUs shared for each sample pairwise was determined using *shared_phylotypes.*py in QIIME1. For each individual, the percentage of an individual's OTUs present in each of the created mixture samples was determined. The percentage of OTUs shared was determined for each of the following categories: inter-individual, intra-couple (individual to the corresponding

couple mixtures), inter-couple (individual to the other couples' mixtures), and each contribution level (1-99%) (individual to the mixture containing their contribution level). These categories were compared pairwise using Welch's Two Sample t-test. Additionally, the mean, maximum, minimum, and standard deviation of each category was calculated. All calculations were performed using R. In order to expand these analyses beyond *in silico* mixtures, the original population Q1 OTU and Deblur OTU data from Chapter 3 were used to repeat the OTU sharing analysis. Samples were first collapsed by individual and collection week by summing in order to capture all of the replicate hair and swab samples in a single sample.

As a proof of concept, the Q1 OTU assignments for the *in silico* mixtures and original population samples were used to evaluate SourceTracker [3] for the prediction of contributors in a sample using a Bayesian framework. SourceTracker is a package available for R and was run following the steps in the included example script ("example.r"). In SourceTracker, individual samples are labeled as "source" or "sink." "Source" samples are used to train the SourceTracker model with "environment" identifying groups of "source" samples with a common origin. "Sink" samples are the samples with unknown contributions from the "source" environments and/or an unknown environment. For the *in silico* mixtures, the samples which contained 100% of an individual were set as "source" with the individual label as the "environment." The mixture samples were limited to pseudo-Couple 1 and were set as "sink" with the environment as "mixture." The

alpha1 and alpha2 values were set at 0.001. These values may be tuned using the sample data for smoother distributions for low-coverage samples. Here, all samples contained 10,000 reads. For SourceTracker training and source proportion estimates, the samples were rarefied to 5000 reads. Original population samples were collapsed on participant code and collection week by summing, combining hair and swab replicate samples for each individual at each time point. Collapsing the samples in this fashion resulted in greater than 28,000 reads per sample. For each individual who was not sexually active in the week prior to all of their sample collections or mixed in their activity, the first time point sample associated with an inactive week was selected as the “source” sample and the environment labeled with the participant code. If an individual was sexually active in the week prior to all of their sample collections, the sample with the largest time interval prior to sample collection was selected as the “source” sample and the environment labeled with the participant code. The remaining samples were set as “sink” samples with the environment label “mixture.” SourceTracker training and creation of source proportion estimates were performed as before with a rarefaction level of 5000 reads.

4.4 Results

4.4.1 Shifts in Beta Diversity

Within each “pseudo-couple” mixture set there is a clear shift in beta diversity from one individual to the next as the proportionate contribution of the donor increases, using the weighted UniFrac. Figure 4.1 illustrates how PC1 through PC3 vary continuously between the two individuals in each couple. The unmixed individuals

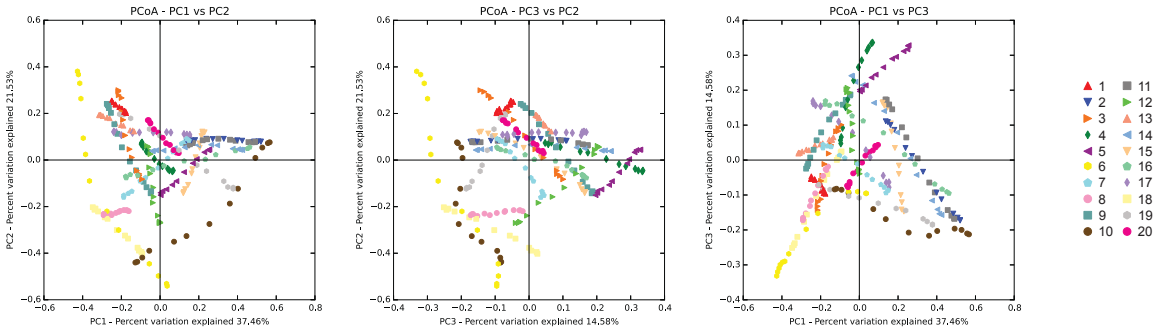


Figure 4.1 Weighted UniFrac PCoA plot of *in silico* couple mixtures.

are at either end of each line of samples representing a couple. The shift in beta diversity with changes in mixture ratios follows a regular pattern with smaller changes in mixture ratios resulting in smaller shifts on the principal components plot. Because the weighted UniFrac takes the relative abundances of the OTUs into account, this regular patterning of the PCoA plots is expected. Within couples, some couples are more similar (pseudo-Couple 8) to each other while others are

more diverse (pseudo-Couple 6). The ability to visually group a mixed sample with its couple depends both on the similarity of the couple and how similar that couple is to other couples. Pseudo-couple 5 and pseudo-Couple 8 are distinct from each other across all PC1-PC3 such that a mixed sample of one of those couples may be confidently assigned. Meanwhile pseudo-Couples 2 and 11 overlap across all axes and would be indistinguishable. When looking all of the couples together, the shifts in beta diversity with shifts in mixture ratios is still visible with *a priori* knowledge of the couples but would it would be impossible to assign a mixture sample to a couple without knowledge of the individual microbiomes before mixing. Presumably, as the total pool of potential contributors to a mixture increases, the probability that any one pair of individuals generate uniquely identifiable mixtures becomes too small to be of practical utility. However, these plots are a gross comparison of shifts in beta diversity and other methods may reveal better predictions of mixture contents.

4.4.2 Percent Mixing Required to Infer Transfer

In order for transfer of microbiome from one pubic region to another to be detectable, it is necessary for a minimum percentage sharing of the donor OTUs to constitute a proportion of the putative mixture of OTUs from a host that leads to greater diversity than would be expected by repeat sampling of the host alone. I estimated this proportion by generating *in* silico mixtures of microbiome profiles from two randomly chosen opposite-sex individuals from the dataset, repeating this process 20 times to generate a representative frequency distribution of

mixtures. Once for each pair, I mixed samples in the proportions 1%, 5%, 10%, 25%, 33%, 50%, 66%, 75%, 90%, 95%, and 99% by randomly sampling the appropriate number of reads from host and donor to a total of 10,000 reads. Rather than simply mixing the proportions in the two individuals, this process was intended to simulate stochastic transfer of taxa. Two measures of taxon sharing were assessed, namely the proportion of shared OTUs and the weighted UniFrac distances between samples.

Applying the OTU sharing measure to the Q1 taxon assignments suggests that a minimum of 10% donor contribution is required to observe a significant shift in the density distribution of the percent OTUs shared between a mixture and a pure donor sample (Figure 4.2a). Typical individuals share a mean of 59% of their Q1 OTU, albeit with a wide range from 25% to 80%; the distribution of sharing is indistinguishable from that measured between an individual and 1% contribution sample. Reliable separation is not achieved until at least 33% mixing, at which point the mean proportion of sharing has increased to 82% (Table 4.1). Thereafter, additional contributions from the donor result in increased mean and minimum proportions of sharing between the mixture and donor. It is thought unlikely that in a forensic situation the victim's microbiome would contribute less than half of the swab or sample, so the blue-green hued curves to the right of the panel simply confirm that strong mixing results in the expected similarity to the donor.

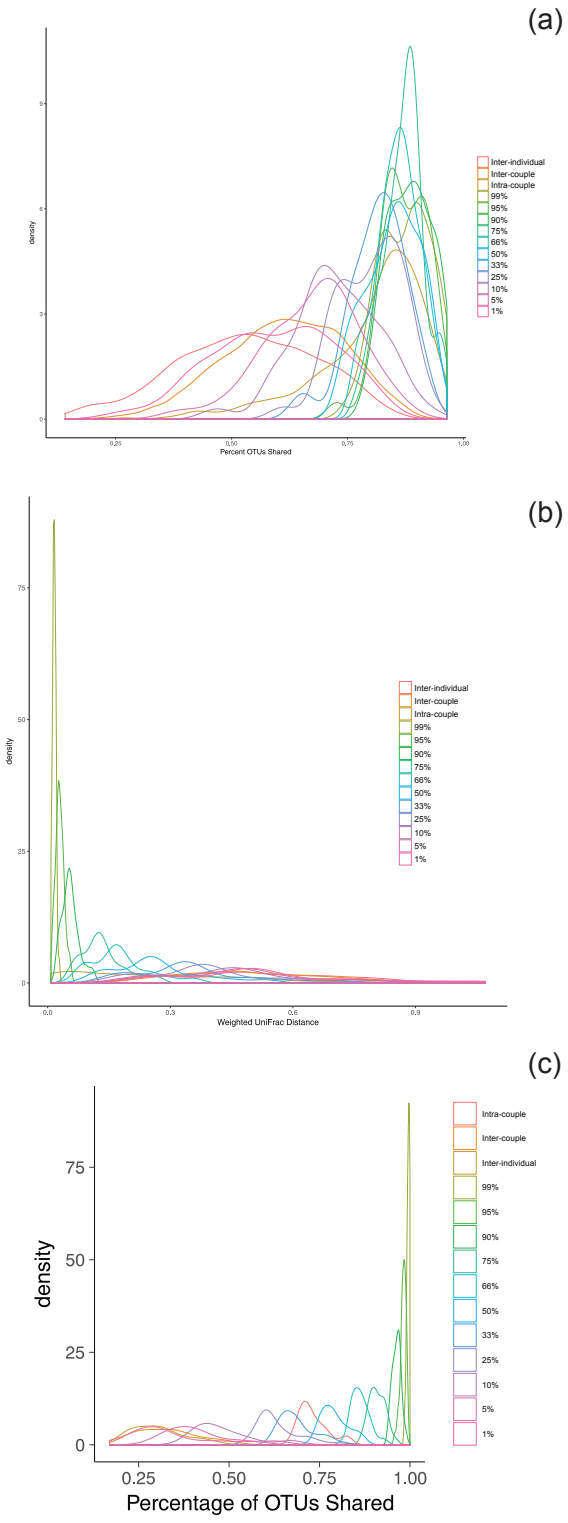


Figure 4.2 Density plots of various categorical groupings of *in silico* mixtures by (a) percentage of Q1 OTUs shared, (b) weighted UniFrac distance, and (c) percentage of Deblur OTUs shared.

Table 4.1 Maximum, minimum, mean and standard deviation for (a) Q1 percent OTU sharing, (b) weighted UniFrac distance, and (c) Deblur percent OTU sharing for *in silico* mixtures.

(a)	Q1 Percent OTU Sharing			
Category	Maximim	Minimum	Mean	Standard Deviation
Inter-Individual	0.87	0.14	0.53	0.15
Inter-Couple	0.95	0.16	0.61	0.13
Intra-Couple	0.96	0.27	0.80	0.12
99%	0.96	0.75	0.88	0.05
95%	0.96	0.73	0.87	0.05
90%	0.96	0.80	0.88	0.05
75%	0.96	0.77	0.87	0.04
66%	0.95	0.77	0.86	0.05
50%	0.93	0.74	0.85	0.06
33%	0.92	0.65	0.82	0.06
25%	0.93	0.61	0.80	0.07
10%	0.92	0.47	0.72	0.09
5%	0.86	0.39	0.67	0.10
1%	0.80	0.27	0.59	0.13

(b)	Weighted UniFrac Distance			
Category	Maximum	Minimum	Mean	Standard Deviation
Inter-Individual	1.071	0.103	0.522	0.191
Inter-Couple	1.065	0.053	0.494	0.179
Intra-Couple	1.057	0.008	0.266	0.227
99%	0.028	0.008	0.016	0.004
95%	0.055	0.012	0.030	0.010
90%	0.112	0.025	0.056	0.021
75%	0.269	0.060	0.134	0.053
66%	0.363	0.076	0.177	0.072
50%	0.534	0.110	0.265	0.107
33%	0.715	0.142	0.353	0.143
25%	0.802	0.153	0.397	0.162
10%	0.962	0.189	0.476	0.194
5%	1.016	0.213	0.504	0.204
1%	1.057	0.212	0.523	0.213

(c)	Deblur Percent OTU Sharing			
Category	Maximum	Minimum	Mean	Standard Deviation
Inter-Individual	0.78	0.07	0.28	0.10
Inter-Couple	0.78	0.08	0.33	0.10
Intra-Couple	1.00	0.17	0.73	0.24
99%	1.00	0.98	1.00	0.00
95%	0.99	0.96	0.98	0.01
90%	0.99	0.94	0.96	0.01
75%	0.96	0.87	0.91	0.02
66%	0.93	0.81	0.86	0.03
50%	0.87	0.73	0.79	0.04
33%	0.82	0.62	0.69	0.05
25%	0.81	0.55	0.63	0.06
10%	0.68	0.36	0.48	0.08
5%	0.65	0.28	0.42	0.10
1%	0.61	0.17	0.33	0.11

Two analytical refinements increase the ability to detect mixtures. Weighted UniFrac distance distributions in Figure 4.2b have the inverse relationship where small values correspond to high sharing. The distance comparisons are not significantly different from random between-individual contrasts for the 1-10% contribution levels ($p > 0.01$) but are significant for all other comparisons. Even more discrimination appears to be achievable with Deblur OTUs as shown in Figure 4.2c where even 5% mixing generated OTU sharing distributions significantly higher than for inter-individual comparisons, and by 25% mixing, the majority of mixture-donor distances are greater than individual-donor ones. Even though the absolute percentages of sharing are lower than with the Q1 OTUs, the increased diversity improves resolution of mixture proportions. Interestingly, within-couple comparisons fall in-between the 33% and 50% contribution distributions, consistent with cohabitation resulting in this level of Deblur OTU sharing (Table 4.1).

Taken together, these analyses imply that a minimum of 5% mixing is required from microbiome transfer during sexual activity to be able to detect any change in the host's microbiome. However, this proportion needs to be over 25% to provide reliable detection.

4.4.3 Suspect Identification from the General Population

Next, I asked whether, given that transfer has occurred resulting in a mixture different from that of the victim's normal profile, is it possible to infer the identity of

the perpetrator? Random Forest analyses applied to the 260 *in silico* mixtures correctly identified a member of that pseudo-couple almost 50% of the time (129/260), predominantly when the identified individual was the major contributor to the mixture. In the case of a transfer of the pubic hair/pubic area microbiome, it is likely that the major contributor on a sample will be the host, i.e., the victim if it is the victim's pubic hair. Of the incorrect matches, 119 were due to just seven individuals whose profiles are intermediate enough within the entire sample to have an elevated probability of matching many other individuals. I also asked whether the Random Forest matching may be due to, or adversely affected by, the highest abundance taxa, but found no evidence that is the case as they were not enriched among the inferred "high importance" OTU. Removal of the high importance taxa also had little impact on the proportion of couple matching.

SourceTracker was used to evaluate how well this tool would work with forensic-type samples. SourceTracker was originally developed for tracking the source of laboratory contaminants from known and unknown sources. In that implementation, multiple source samples were from larger environments such as fecal, oral, and soil sources where each microbiome environment is well separated and easily distinguishable. Here, the source samples are from one location, the pubic region, with each individual set as a different "environment" modeled in the SourceTracker framework. I also note that only one sample from each individual was used as a source for training the SourceTracker model such that intra-individual variability was not captured.

To initiate testing of SourceTracker with the pubic hair/pubic swab samples, mixture samples from one pseudo-couple (pseudo-Couple 1) were used as the unknown ("sink") samples with the pseudo-Couple 1 single-source samples used as the known environment ("source") samples. The predicted identities and proportions of contributors were similar to the actual input amounts (Figure 4.3) with small proportions predicted to be from an unknown source. Given the stochastic nature of the sampling and OTU sharing, it is to be expected that the predicted mixture ratios would not exactly match the *a priori* mixture ratio. Encouragingly, the unknown contribution was not predicted to be more than 1.4% for any of the mixtures.

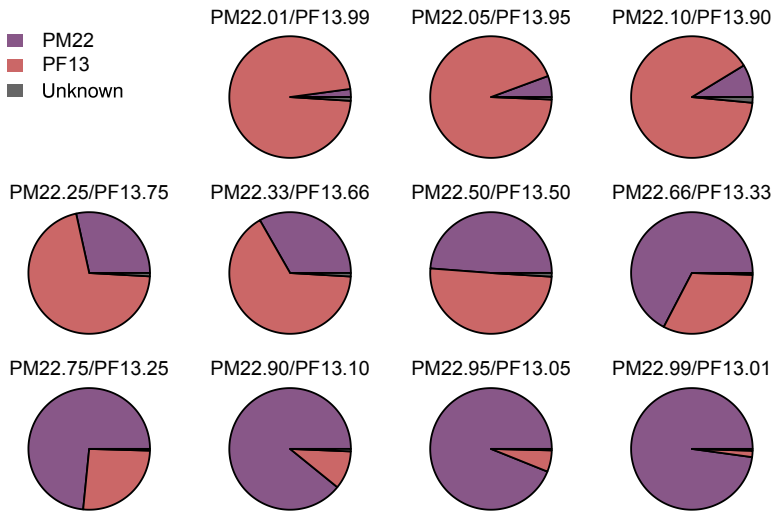


Figure 4.3 Predicted sources to the pseudo-Couple 1 mixtures using the individuals in pseudo-Couple 1 as the source environments.

In order to test the effects on the predictions if one contributor is not available as a source sample, the analysis was repeated by training the SourceTracker model with one pseudo-Couple 1 contributor source (PF13) and one non-contributor source (PM92). In this case, the predicted contributor input ratios for PF13 roughly followed the known input ratios with PM92 not included as a significant contributor (Figure 4.4). The remaining contributor was labeled as unknown. In this case, the non-contributor individual is not represented by more than 0.73% in any of the mixtures. The unknown source environment proportion is larger than when both

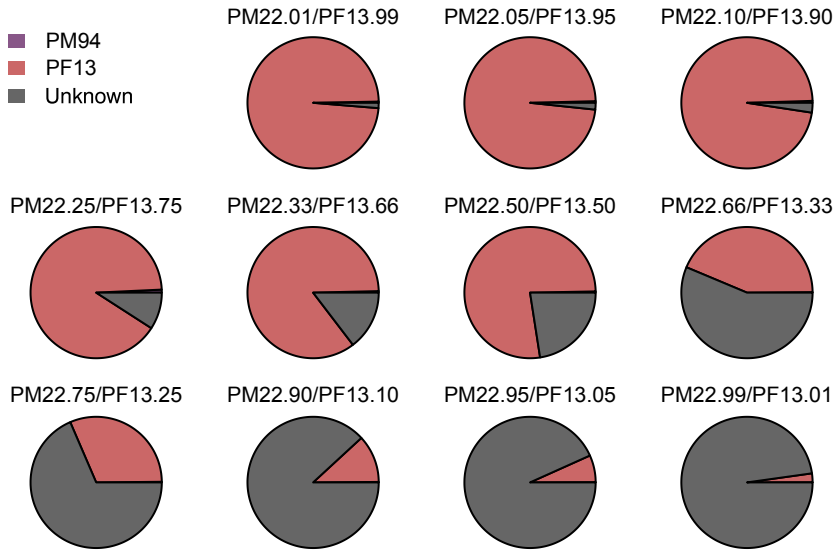


Figure 4.4 Predicted sources to the pseudo-Couple 1 mixtures using one individual in pseudo-Couple 1 and a non-contributing individual as source environments.

individuals in the couple were source environments but that proportion is not equal to the sum of the original unknown proportion plus the missing source environment individual.

The pool of source environments was expanded to include all 40 of the individuals used to create the *in silico* mixtures in order to explore the effects a larger pool of source environments would have on the SourceTracker predictions. This analysis represented ten times the number of source environments for training the SourceTracker model than was originally reported with the development of the tool. Here, the correct individuals were predicted at ratios similar to the known input amounts (Figure 4.5). The non-contributing individuals and unknown source

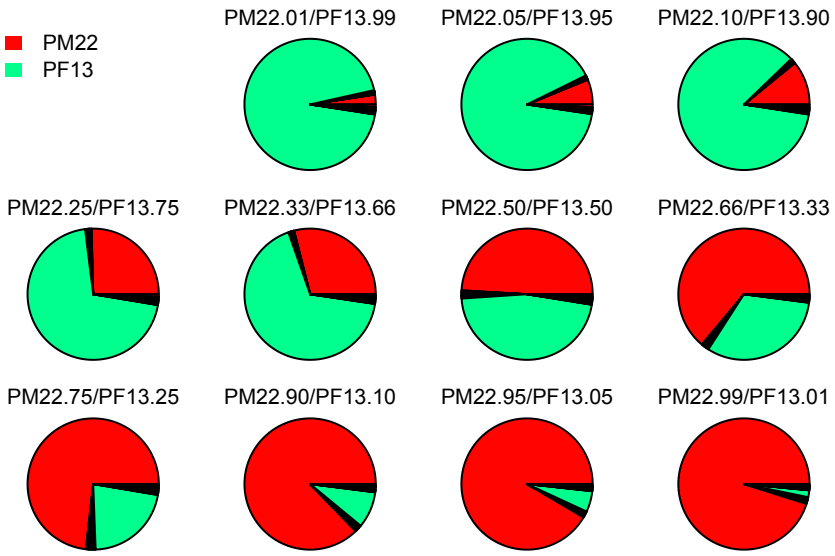


Figure 4.5 Predicted sources to the pseudo-Couple 1 mixtures using all individuals as source environments.

contributors represent 3.2-5.0% to the mixtures. These results reflect the increased chance of similarity with an unrelated individual as the pool of source environment individuals increases.

Given the promising results with these *in silico* mixtures for the prediction of proportions of potential source environments, the SourceTracker analysis was extended to the original population data and couples. Again, only one sample per individual was used to establish a source environment for that individual as opposed to multiple samples from an individual capturing intra-individual variation for the model. Many of the samples were predicted to have multiple (>2) source individuals contributing to the mixture (Figure 4.6). Of the 115 samples identified by SourceTracker as unknown mixtures, 37 (32%) had the correct individual assigned as the largest contributor to the sample. Two couple samples had their partner assigned as the second largest contributor and two couple samples had their partner assigned as the largest contributor. The proportion of contribution of the correct donor of the sample ranged from 0.08 to 91% with a mean of 25% $\pm$ 26%. An unknown contributor was the largest or second-largest contributor in 24 samples. In 26 out of 38 (68%) samples collected with no sexual activity in the week prior to collection and in 48 out of 71 (68%) samples collected with sexual activity in the week prior to collection, the correct individual was not the largest contributor to the sample. These results potentially reflect the effects of intra-individual variation over time leading to more inter-individual similarity along

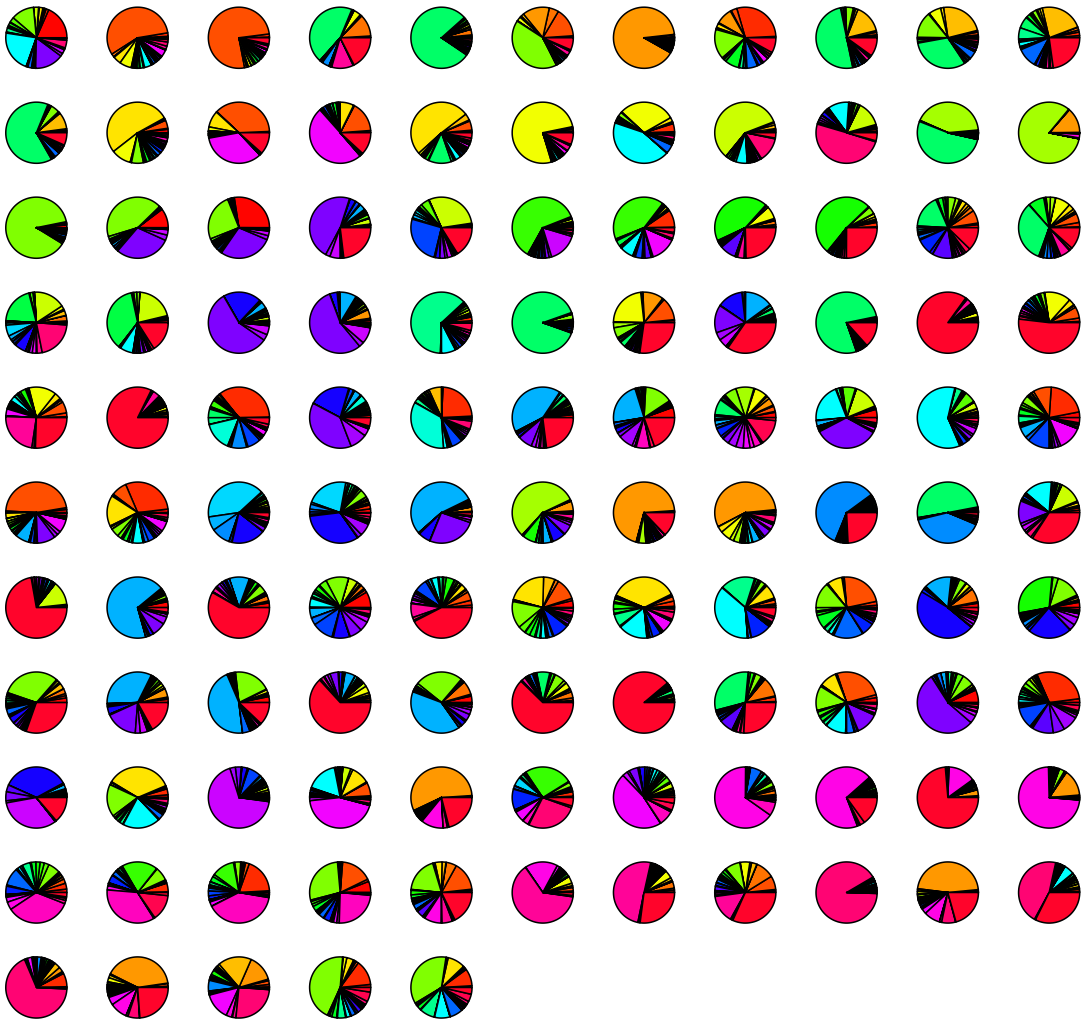


Figure 4.6 Variation in predicted contributor proportions in population samples as determined by SourceTracker.

with increased probability of matching an unrelated individual with increasing population pool. Additionally, these sample represent closely-related environments with a higher degree of similarity than when comparing environments such as fecal, oral, and soil samples. Using longitudinal samples from individuals to create a larger source environment for each individual may improve results by accounting for the intra-individual variation. Results may also

be improved by limiting the source environments to a smaller pool of suspects with potential for excluding suspects from being predicted contributors.

4.4.4 Culprit Identification from a Small Pool of Suspects

The above analyses demonstrate that it is highly unlikely that the perpetrator of a sexual assault can be identified beyond reasonable doubt by matching a mixture sample to a database of pubic microbiome samples. Although up to half of the matches in the *in silico* mixtures are "correct," this is when there is a high donor level and is from a small sample of the general population and, as noted previously, as the potential source population increases, the uncertainty of matching will become too great to be useful.

However, an alternative application of forensic microbiome profiling may be to confirm that a suspect, or one of a small set of suspects identified by the victim or other means, was involved in the crime. Situations may also arise where the suspect admits to contact with the victim, but denies that sexual intercourse took place, and pubic microbiome analysis could be used to refute this claim.

Evaluation of this probability depends on the development of a precise diagnostic that has high sensitivity (identifies transfer from the culprit to the victim if it occurred) and high specificity (does not report a false positive commonly). These data are typically presented in the form of a Receiver Operating Characteristic (ROC) curve, such as those in Figure 4.7 for the Q1 and Deblur OTU analysis of

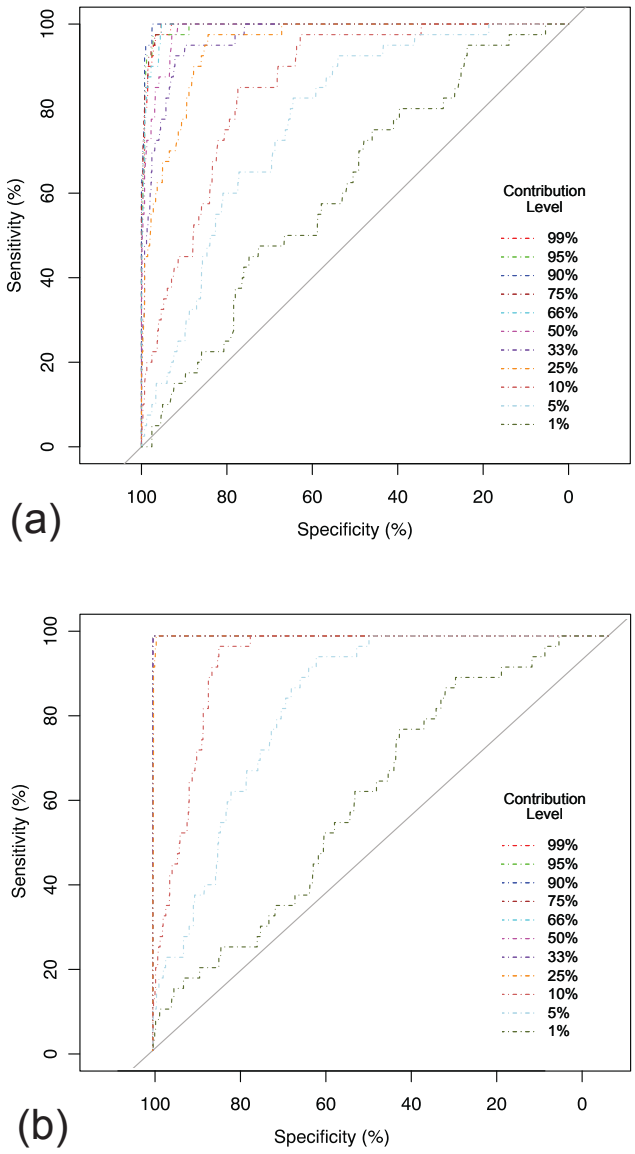


Figure 4.7 Percent OTU sharing ROC curves by contribution level for (a) Q1 OTUs and (b) Deblur OTUs.

this dataset. For forensic purposes, the precision of the test is also critical, which is the proportion of times a positive test is actually due to mixing from the culprit. As shown in Figure 4.8, the precision is influenced by the prior likelihood that the

Equal Prior Probability

	Transfer	No Transfer		
Predicted mixture	85	23	79%	Precision
Predicted not mixture	15	77	84%	NPV
	85% Sensitivity	77% Specificity		

10% Prior Likelihood

	Transfer	No Transfer		
Predicted mixture	85	230	27%	Precision
Predicted not mixture	15	770	98%	NPV
	85% Sensitivity	77% Specificity		

Figure 4.8 Precision and negative predictive value (NPV) of Q1 OTU sharing at 10% contribution level and 65% OTU sharing for equal prior likelihood and 10% prior likelihood.

mixture includes the culprit's microbiome, so has a Bayesian posterior probability interpretation. The data is taken from the 10% contribution threshold in Table 4.2, which summarizes the best threshold observed for discriminating mixtures from individuals for various levels of contributions. At this level of contribution, if 65% OTU sharing is observed between the mixture and a sample from the suspect, there is 85% sensitivity and 77% specificity. Assuming an equal prior probability that the suspect is or is not the culprit, the precision would be close to 80%; but if this likelihood drops to just 10%, then the precision is only 27% and the positive

interpretation of mixing is more likely due to some other explanation (such as similarity of the suspect's microbiome to the culprit's).

Table 4.2 ROC summary for Q1 and Deblur percent shared OTUs.

	Best									
	Threshold		Sensitvity		Specificity		AUC (%)		95% CI (%)	
Contribution	Q1	Deblur	Q1	Deblur	Q1	Deblur	Q1	Deblur	Q1	Deblur
1%	0.518	0.231	75.0	90.0	46.0	33.5	61.3	61.4	53.1-69.5	53.5-69.3
5%	0.594	0.310	82.5	95.0	64.5	64.1	77.6	83.7	71.7-83.5	79.5-87.8
10%	0.653	0.396	85.0	97.5	77.4	85.5	85.8	93.6	81.4-90.2	91.7-95.4
25%	0.697	0.549	97.5	100.0	84.4	99.2	95	99.9	92.9-97.2	99.7-100.0
33%	0.734	0.609	95.0	100.0	89.8	99.9	96.7	99.9	95.0-98.4	99.8-100.0
50%	0.744	0.676	100.0	100.0	91.4	99.9	98.5	99.9	97.7-99.2	99.9-100.0
66%	0.772	0.797	100.0	100.0	95.3	100.0	99.3	100.0	98.8-99.7	100.0-100.0
75%	0.773	0.825	100.0	100.0	95.4	100.0	99.5	100.0	99.2-99.9	100.0-100.0
90%	0.798	0.860	100.0	100.0	97.5	100.0	99.7	100.0	99.5-99.9	100.0-100.0
95%	0.796	0.870	97.5	100.0	97.3	100.0	99.3	100.0	98.8-99.9	100.0-100.0
99%	0.789	0.883	97.5	100.0	96.7	100.0	99.3	100.0	98.9-99.3	100.0-100.0

These probabilities are encouraging for positive confirmation that a suspect contributed to a mixture, and are improved upon substantially if the proportion of mixing is over 33% or if the Deblur OTU are used. If fact, above 25% contribution, the AUC with Deblur is close to 100%, implying high precision unless the prior odds that the suspect is the culprit are very low.

The real couples showed evidence of mixing based on these *in silico* power calculations. Additionally, there are multiple samples over multiple time points for each individual, capturing some of the natural variation that might be observed in

individuals and couple mixtures. The overall distribution of percentage of OTUs shared are quite similar for the intra-individual (Self), intra-couple (Couple), and inter-individual (Unrelated) categories (Figure 4.9) with more separation between

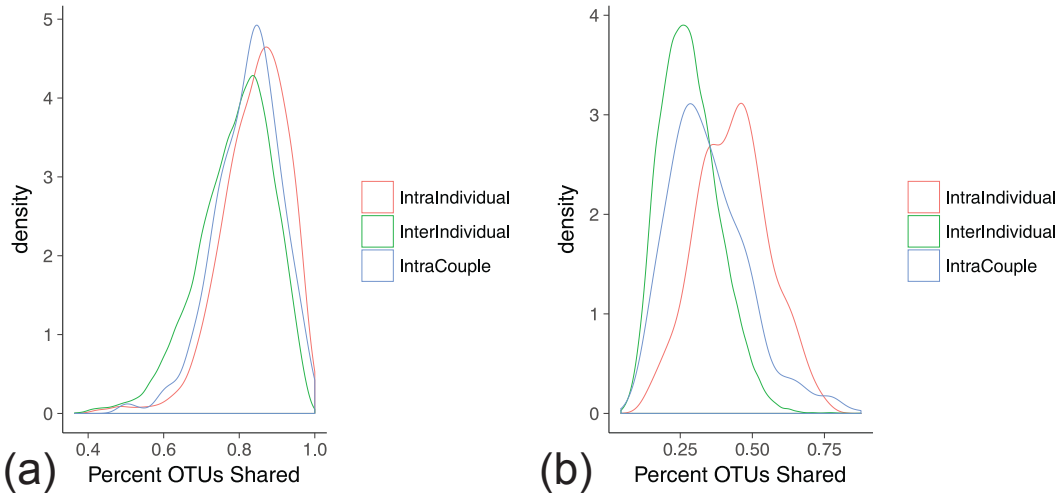


Figure 4.9 Density plot of percentage of OTUs shared for the categorical groupings of population samples using (a) Q1 generated OTUs and (b) Deblur generated OTUs.

the categories using the Deblur-generated OTUs. The differences in distributions are significant using either method (Q1 or Deblur, respectively) of generating OTUs for all three comparisons of Self-Couples ($p = 6.4\text{x}10^{-4}$, $1.8\text{x}10^{-24}$), Self-Unrelated ($p = 2.1\text{x}10^{-29}$, $4.2\text{x}10^{-88}$), and Couples-Unrelated ($p = 1.4\text{x}10^{-17}$, $2.14\text{x}10^{-25}$), with the expected trend of increased OTU sharing from Unrelated to Couples to Self. The distributions are broader than in the *in silico* data, reflecting stochastic sampling and possibly also variability in mixture proportions for couples

across time. All categories exhibit a maximum OTU shared percentage of 100% for Q1 OTUs and higher maximum OTU shared percentage for inter-individual samples than for intra-individual samples using Deblur OTUs (Table 4.3). Because the maximum OTU shared percentage is the same for all categories, the ability to distinguish mixed samples from unrelated individuals is diminished. The Deblur OTUs showed overall lower levels of OTU sharing as would be expected with the finer classification level achieved with this method. Analysis is also complicated by not knowing the contribution level of each individual in a sample in order to select the correct ROC curve for power calculations.

Table 4.3 Means, maximums, minimums, and standard deviations of percentage of OTUs shared in population samples.

	Maximum		Minimum		Mean		Standard Deviation	
Category	Q1	Deblur	Q1	Deblur	Q1	Deblur	Q1	Deblur
Intra-individual	1.000	0.754	0.429	0.141	0.841	0.432	0.087	0.124
Inter-individual	1.000	0.877	0.365	0.044	0.790	0.284	0.099	0.101
Intra-couple	1.000	0.877	0.492	0.086	0.822	0.346	0.086	0.138

4.4.5 Exculpation of Falsely Accused from a Small Pool of Suspects

The reciprocal situation arises where a suspect seeks exculpation for a crime he did not commit. If there is no detectable transfer, any microbiome information is uninformative. However, in the event that it can be established that a sample taken soon after the alleged crime is unlike the victim's pre-assault sample or

unlike her longitudinally established profile, there may be prospects for excluding one or suspects as the culprit. Possible situations include multiple suspects who may have been present at the time of the crime and date-rape, where there is an established history of relationship between the two individuals that does not include intercourse.

The inverse of precision is negative predictive value (NPV), namely the proportion of situations where the sample is predicted not to have contributed to a mixture and actually did not. In Figure 4.8 the NPV are 84% and 98% for the equal prior and 1:10 ratio, respectively. Perhaps more commonly, though, will be situation where the suspect seeking exculpation will be one of several, and the challenges is to evaluate the posterior probability for each person. In this situation, there should be high power to establish that the likelihood of a mixed microbiome profile given that one suspect is the culprit is significantly lower than that of one or more of the others, providing significant doubt.

4.5 Discussion

The microbiome is a complex ecosystem with many analytical challenges even when considering it originates from a single source, so allowing for mixtures of microbiomes increases the complexity of the analyses. There have been multiple studies considering the fate of introduced bacteria or microbiomes. In some of

these cases, the introduced microbiome is from a different body site (oral to skin) [9] or is a marker bacteria not commensal to the recipient site [2]. Tracking of these introduced species is a relatively simple task. In the case of fecal microbiome transplants (FMTs) for the treatment of *C. difficile* infections, the recipient microbiome is highly perturbed by the infection and distinct from a normal gut microbiome. For these studies, the introduced bacteria persisted for varying amounts of time. Oral to skin transfers persisted on the time scale of hours or longer depending on whether the skin site was sebaceous [9]. In the same study, skin to skin transfers (forehead to forearm) were immediately more similar to their initial state. For oral microbiomes, couples who intimately kissed frequently were found to have more similar profiles than those couples who kissed less frequently [2]. When one partner was inoculated with a probiotic yogurt drink, levels of the yogurt bacteria as measured by percent of total bacterial population dropped sharply between the first partner and the second after kissing. Samples were collected immediately after inoculation or kissing, so it would be expected that the yogurt bacteria levels would continue to drop over time. In the case of FMT, donor bacteria have been detected using shotgun metagenomics months after the transplant, but primarily as different strains of shared species as opposed to novel species introduced through the transplant [10].

Here, I am interested in the detectible transfer of bacteria between the same body site of two individuals. This analysis is complicated by the similarity in bacterial composition between individuals at a particular body site. Additionally, how much

bacteria are transferred plays a role in the detection of the transfer. While the *in silico* mixtures represented the best-case scenario where exact contributor levels were known, the original population data revealed how much more variability there was in the samples. Given this variability and unknown contributor levels within *in vivo* samples, there may be more value in the exculpatory power of these analyses than in inclusion of a suspect in a purported mixed sample.

For discrimination of sample, the Deblur OTUs outperformed the Q1 OTUs. The total number of OTUs in the data set went from 307 OTUs to 8913 OTUs when using Deblur. Overall, the percentage of OTU sharing was lower when using Deblur but would potentially be more indicative of related samples when OTUs are shared. In comparing a suspected mixture sample (i.e., a pubic hair sample from a victim) to a reference sample (i.e., a suspect in a forensic case), the percentage of OTU sharing was quite sensitive and specific when there was at least 79% OTU sharing and at least 66% contribution from the reference. Between the 25% and 50% contribution levels, percent OTU sharing at a threshold of 55-68% also had over 99% specificity. The maximum percentage of OTU sharing for inter-individual comparisons was 78% while the minimum percentage for a 50% contributor was 73%, indicating that a result of 73% OTU sharing or above would be due to the presence of the referenced individual at 50% or more.

Given that a contribution of at least 25% was needed to provide a reliable detection of a mixture, the question of whether this much donor microbiome is likely to be transferred to the host is important and needs to be resolved. This question can also be broken down between the hair and swab samples. Overall, the swab samples exhibited a higher level of alpha diversity, meaning more OTUs were detected on the swab samples than the hair samples. Higher specificity and sensitivity to detect mixtures was demonstrated using the Deblur OTUs where there were more than 10 times the OTUs than with the Q1 OTUs. Therefore, more consistent results may be attainable by focusing on pubic swabs with Deblur OTU assignments.

Application of a Bayesian approach through the SourceTracker analysis also supported the potential to discriminate contributors to a sample to include unknown contributors. The initial suggested use of SourceTracker was to determine sources of potential contamination [3]. In this case, the sources varied from the gut, oral, and skin to soil. These sources have quite distinct microbiomes and the sources used for training the SourceTracker model consisted of multiple samples, capturing the variation within a given environment (i.e., gut or soil). Here, SourceTracker was asked to discriminate between individuals where the microbiomes are from a single environment type with much lower variation between individuals. The power of discrimination needs to be much better in order to distinguish more subtle differences. Using the mixture samples from one *in*

silico couple, SourceTracker was able to recapitulate the rough input ratios of the couple at all levels when given up to 40 reference individuals.

Extension of these analyses to the larger population data set reveals the differences between couples and unrelated individuals are harder to discern. The contribution levels of any individual to the samples are unknown. Multiple samples from each individual are also included, introducing intra-individual variation not present in the *in silico* samples. Some inter-individual percent OTU sharing was higher than the intra-individual percent OTU sharing. The minimum percent OTU sharing for the intra-individual comparison was 14% and 4.4% for the inter-individual comparison. Thus, a lower percent OTU sharing threshold may be set in order to state that a suspected individual has the potential to be excluded from an evidence sample. Comparing percent OTU sharing within smaller subsets of individuals may be more discriminating with higher exculpatory value. Given that the pool of suspects for a given sexual assault is limited, being able to include or exclude a particular suspect as compared to the entire male population would not be necessary.

SourceTracker also resulted in more varied responses. The role of intra-individual variation appears to play a role in these responses as shown by the 68% of samples with no sexual activity in the week prior to sample collection not having the correct individual as the primary contributor to the sample. Further refinements to the model may improve results. Given the variation in an individual over time, it

may be necessary to compare an evidence sample to a reference sample taken from the victim and suspect within a short time frame from the incident, though there may still be a mixture present on the sample. This mixture on a suspect reference may improve the predicted proportion of this individual to the evidence sample as it would be more similar to the evidence sample. The *in silico* mixtures suggest that the inclusion of a suspect of more than 5% by SourceTracker means that it is likely that that suspect is an actual contributor to that sample. As evidenced by the population samples, SourceTracker predicted contributors above 5% who were known not to have contributed to the samples. Again, there may be more power of exclusion with this tool when extended to the larger population data set and with a smaller pool of credible suspects set as source environments.

How to best use microbiome data may be a matter of what question is being asked. If the question being asked is one of identity of the donor of a pubic hair left at a scene, it may be sufficient to compare the microbiome developed to a suspect and then to a database for weighting of the match. If the question under consideration is the detection of mixtures of pubic hair/area microbiomes after sexual contact, it may be more informative to collect standards from both individuals temporally close enough to the event to minimize normal variation but long enough from the event to allow for the individuals' microbiomes to return to a non-mixture state. These standards may then be compared to pubic hairs/swabs collected directly from an individual. Based on the Deblur *in silico* mixtures,

sharing of OTUs over 67% between the evidence hair/swab and non-owner individual would be suggestive of the presence of at least 50% that individual in the sample. Reshuffling of the couples and repeating the rarefactions could be performed so that more population variation may be captured to explore how this variation affects the models. Additionally, *in silico* simulations could be done where the mixtures are made from one time point's sample while a different time point sample is used as the reference to model the temporal variation. Studies with controlled sample collections both before sexual activity and collections at set time intervals after sexual activity will provide insight into how much of an individual's microbiome is transferred during sexual activity and how long the transfer can be detected/how long it takes for the pubic hair microbiome to return to normal after sexual activity. These efforts could be used to create Bayesian models and provide priors for the model. The potential use of Bayesian models is discussed further in the following chapter.

It may be necessary to use shotgun metagenomics, which can detect bacteria at the strain level where targeted 16S sequencing classifies at the genus and some species level, to detect these pubic hair/pubic area microbiome transfers. The Deblur data supports the use of analyses that can provide this finer resolution of OTUs. The utility of shotgun metagenomics has been demonstrated in the monitoring of fecal microbiome transplants [10]. The therapeutic transfer of the fecal microbiome mirrors that of the pubic hair/area microbiome in that the transfer of like-to-like communities. However, FMT involves the transfer of a large amount

of donor material. Even so, researchers were able to differentiate bacteria from the donor and recipient that were the same at the genus/species level but differed at the strain level. This level of differentiation may hold promise for the forensic comparisons similar to the promise of single nucleotide polymorphism detection within forensic human STRs using next generation sequencing for mixture deconvolution.

The mixture modeling results here show promise for the construction of models and methods for detecting and interpreting microbiome mixtures. Further work with controlled sampling for refining the models will continue this technique forward. Analysis of the pubic hair/pubic area microbiome for mixtures has the potential to assist with the investigation of sexual assaults that further study will improve.

4.6 References

[1] Tridico S, Murray DC, Addison J, Kirkbride KP, Bunce M. Metagenomic analyses of bacteria on human hairs: a qualitative assessment for applications in forensic science. Investigative Genetics. 2014;5.

[2] Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. Microbiome. 2014;2:41.

[3] Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nature Methods. 2011;8:761-5.

[4] Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nature Methods. 2010;7:334-6.

[5] Amir A, McDonald D, Navas-Molina JA, Kopylova E, Morton JT, Xu ZZ, et al. Deblur rapidly resolves single-nucleotide community sequence patterns. eSystems. 2017;2:e00191-16.

[6] Lozupone C, Knight R. UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. Appl Environ Microbiol. 2005;71:8228-35.

[7] Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R. UniFrac: an effective distance metric for microbial community comparison. The ISME Journal. 2011;5:168-72.

[8] Liaw A, Wiener M. Classification and Regression by randomForest. R News. 2002;2:18-22.

[9] Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. Bacterial Community Variation in Human Body Habitats Across Space and Time. Science. 2009;326:1694-7.

[10] Li SS, Zhu A, Benes V, Costea PI, Hercog R, Hildebrand F, et al. Durable coexistence of donor and recipient strains after fecal microbiota transplantation. Science. 2016;352:586-9.

CHAPTER 5

THE PATH FORWARD

5.1 The Pubic Area Microbiome and Forensics

All forensic DNA testing laboratories in the United States that participate in the Combined DNA Index System (CODIS) must follow the Federal Bureau of Investigation's "Quality Assurance Standards for Forensic DNA Testing Laboratories" (QAS) [1]. These standards are meant to ensure that any DNA data being uploaded into CODIS is of sufficient quality and integrity. As CODIS is comprised solely of human short tandem repeat (STR) data at specific loci generated by capillary electrophoresis, other analysis methods and novel markers are not included in these standards. Revisions are currently underway to include next generation sequencing (NGS) methods for forensic human identification in these standards [2]. While there is mounting evidence that an individual's pubic area microbiome is individualizing (Chapters 2-3), it is not yet at the point where it can be used to identify an individual in a crime laboratory setting. Even though the QAS does not apply to the microbiome, we can look to these standards to help guide us in the development of a forensic testing scheme involving the microbiome.

Introduction of a new analysis methodology requires extensive testing of that method, referred to as validation. Different disciplines (i.e., DNA, trace evidence, latent prints, etc.) have different validation requirements. Validation of new forensic DNA methods is covered by Standard 8 of the QAS. Validation is divided into developmental validation and internal validation. Developmental validation covers the testing of a new or novel technique or method prior to its introduction into a forensic testing laboratory. Typically, developmental validation is performed by private companies in the development of commercial off-the-shelf forensic testing kits as public forensic laboratories do not have the time or resources to originate novel DNA methodologies. Internal validation involves the testing of established methodologies within a particular laboratory to show that the methodology works within that laboratory and what the limitations of the methodology are within that laboratory. Looking at the developmental validation requirements set forth by the QAS (Standard 8.2.1) gives a framework for the development of a forensic microbiome analysis methodology.

Standard 8.2.1 of the QAS states:

“Developmental validation studies shall include, where applicable, characterization of the genetic marker, species specificity, sensitivity studies, stability studies, reproducibility, case-type samples, population studies, mixture studies, precision and accuracy studies, and PCR-based studies. PCR-based studies include reaction conditions, assessment of differential and preferential

amplification, effects of multiplexing, assessment of appropriate controls, and product detection studies. All validation studies shall be documented." [1]

A discussion of all of issues that must be addressed before the microbiome, in general, and the pubic hair/area microbiome, specifically, may be used in the crime laboratory are beyond the scope of this work. Many of these issues have been discussed by Clarke et al. [3] and Budowle et al. [4]. A few of these issues are addressed here as it relates to work done in this study.

5.2 Processing Low Biomass Samples and Stochastic Effects

A limitation of many forensic samples is the low amount of biomass available for processing. A consequence of limited sample size is diminution of the accuracy with which a sample represents a given microbiome community. In order to produce meaningful information, the analysis method should be able to produce usable results (i.e., a minimum number of reads above a defined quality level) with low input amounts of DNA. This limitation is even more impactful when considering whole genome sequencing where greater depth of coverage of DNA results in more comprehensive representation of polymorphism the genome. Whole genome sequencing may provide better species and strain-level identifications, but requires larger levels of DNA input than 16S rRNA targeted amplification. Many library preparation kits require a minimum amount of input

DNA, at least 1 ng. Successful recovery of the microbiome DNA begins with efficient DNA extraction methods. Such extractions may incorporate chemical and/or physical disruption methods for releasing the DNA into solution. Additionally, disruption methods affect the types of bacterial DNA recovered as some bacteria are resistant to chemical disruption. Studies comparing extraction methods show that physical disruption, such as bead beating, results in higher recovered microbial diversity to include gram-positive bacteria [5, 6]. A method that can efficiently extract both bacterial and human DNA is desirable as forensic samples are often limited and require efficient use of such samples when multiple testing schemes may be utilized. Another consideration is the potential for an extraction method to be automated. Automation reduces variability inherent in manual methods and can help reduce the potential for contamination.

Library preparation methods also impact the outcome of microbiome results [7]. An individual laboratory must decide which method best suits its needs for the types of samples it will be processing. Additional consideration of community-wide needs must also be considered in order for results to be compared between laboratories. For low-biomass library preparation, there are newer library kits available that allow for lower inputs of DNA as well as other modifications that may be employed. In this study, a modification to the hybridization buffers in the Illumina metagenomic library preparation protocol [8] as provided by Quail et al. [9] was employed targeting normalization of the libraries to 50 pM. There were instances of samples falling below the 50 pM threshold, even after concentration

of sample extracts prior to amplification. Samples below this threshold were still able to produce usable sequencing data, here illustrated by the pubic hair/pubic area swab population samples (Figure 5.1). Samples that were normalized to 50 pM exhibited a range of read counts (52-4,204,850) with a mean of 192,149 reads (standard deviation 278,406) and median of 146,350 reads. After removing the

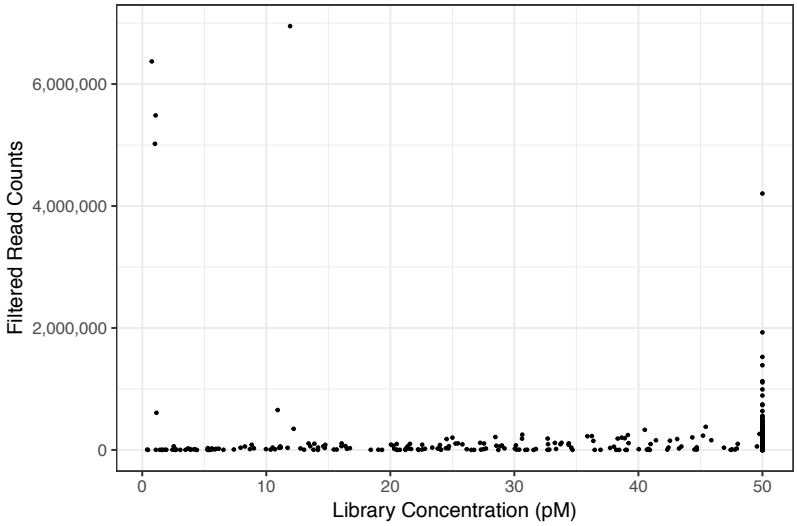


Figure 5.1 Filtered read counts as a function of input sample library concentration.

samples at 50 pM and the samples with greater than 2 million reads from the samples with less than 50 pM, there is a general trend in increasing read counts with increasing library concentration (Figure 5.2).

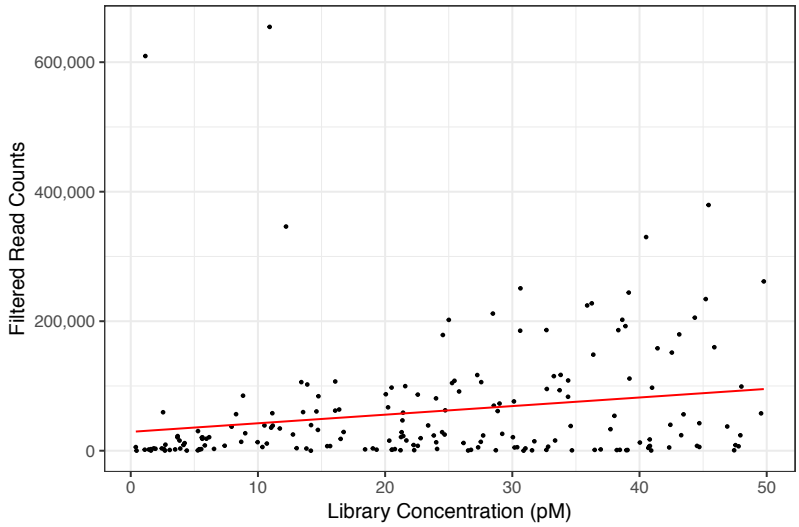


Figure 5.2 Filtered read counts as a function of input sample library concentration with >2 million read outliers at <50 pM removed. Adjusted R^2 = 0.03.

Interestingly, the four highest overall read counts were in samples with less than 15 pM added to pooled libraries, three of which were below 5 pM. These four samples were sequenced in three different pooled libraries such that a single sequencing run was not the cause of elevated read counts for these samples. There was clearly a disconnect between the quantification values for these samples and the actual concentration. Samples that fell out of the upper boundary of the dynamic range (as determined by the C_T of the 20 pM standard) of the quantification kit were diluted and re-quantified using flurometric methods. The same was not done for samples that quantified below the lower boundary of the dynamic range of the quantification kit (0.0002 pM). Improvements to quantification protocols, such as performing qPCR in duplicate on each sample, will be needed to refine the relationship between quantification values and read

counts. Commercial kits for the automatic normalization of libraries are available but typically require larger starting concentrations of DNA than were available here. Despite these challenges, results here demonstrate that developing microbiome profiles is possible from samples with low biomass. The exact lower limit will need to be determined and how best to enhance these samples will require further testing.

5.3 Contamination

A constant concern in the forensic DNA laboratory is the detection and elimination of contamination events. Contamination in human forensic samples refers to the presence of unexpected DNA profiles and is often characterized by a DNA profile in a reagent blank or a DNA profile in an evidence sample that can be linked back to a known individual that is not a victim, suspect, or elimination standard in the case under consideration, such as a laboratory worker. The QAS requires a documented policy for the detection and control of contamination (Standard 9.7). Because of the prevalence of bacteria in the environment, many measures must be used to eliminate the presence of environmental bacteria in and on laboratory equipment and reagents. The typical crime laboratory is not set up to handle bacterial decontamination on the scale as might be found in a microbiology laboratory. A trade-off with an acceptable amount of background bacteria may be required. In this study, reagent blanks and amplification blanks were used as

controls and always produced sequencing reads. The number of control reads were often below the number of sample reads. Rhizobiales, a plant-associated bacterium not commensal in humans, was found in many samples, control and hair/swab, and were removed. Post-Rhizobiales-filtered read counts for reagent blanks and amplification negative controls ranged from 127 reads to 20,794 reads (mean 3590 reads, standard deviation 3946 reads; 65 samples) while the hair and swab samples ranged from 1556 reads to 7,084,537 reads (mean 217,659 reads, standard deviation 571,873 reads; 536 samples). Comparison of the hair and swab samples on a PCoA plot showed that weighted UniFrac diversity was affected by the rarefaction read depth used. At the depth used for analyses of the hair and swab samples in this study (4428 reads), the control samples tend to be clustered together, but were not wholly distinct (Figure 5.3 a) This read depth resulted in filtering out 49 (75%) of the control samples and 11 (2%) of the hair/swab samples. The samples were also rarefied at 1000 and 500 reads (Figure 5.3 b and c, respectively), resulting in 17 (26%) and 8 (12%) of the control samples and none of the hair/swab samples being filtered from the data set. The control samples become even more distinct from the hair and swab sample at lower rarefaction levels, but use of higher read depths for rarefaction is desirable so that more of the variation between samples may be captured. There is a balance between high enough read depth to capture this diversity while low enough read depth to include more samples with varying read depths. The crime laboratory may find it acceptable to perform this filtering so as to confirm the distinct nature of the controls from the evidentiary samples. Or the crime

laboratory may determine what measures it needs to take in order to remove the background level of bacterial contamination.

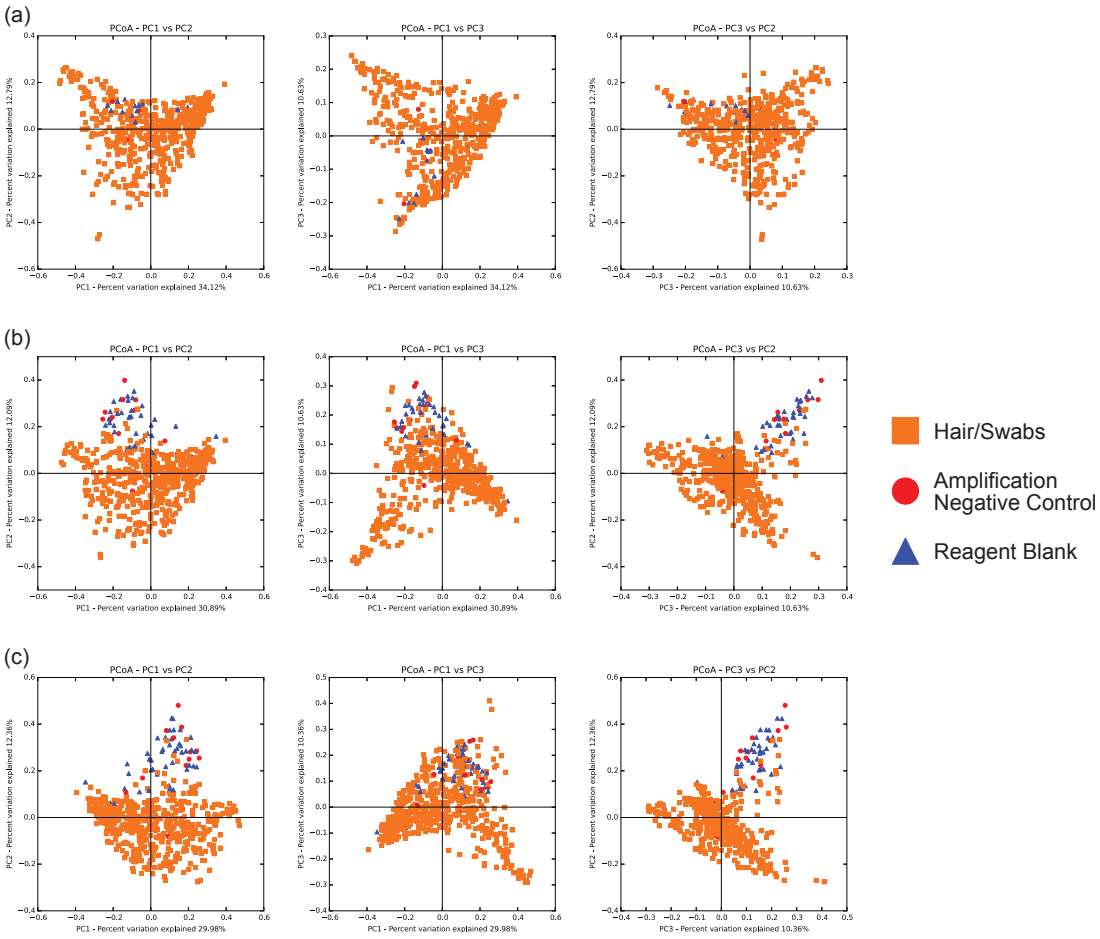


Figure 5.3 PCoA plots of the weighted UniFrac distances at (a) 4428, (b) 1000, and (c) 500 read rarefaction.

5.4 Mixture Analysis

Analysis of mixtures of microbiomes is a challenge, but there is potential to obtain useful information depending on the question being asked. Mixture studies are part of the developmental and internal validation process as proscribed by the QAS (Standards 8.2.1 and 8.3.1, respectively). For human STR mixture analysis, the forensic community is moving towards probabilistic genotyping which utilizes likelihood ratios and some use Bayesian approaches for the interpretation of DNA mixtures. Several software packages are available for this analysis, including STRMix™ (Institute of Environmental Science and Research), TrueAllele® (Cybergenetics), and BulletProof (eDNA). STRmix models behavior such as allelic stutter, degradation, drop-in, and drop-out to model potential mixture combinations and provide likelihood ratios [10]. Bayesian concepts along with likelihood ratios may be used with microbiome mixtures also.

Various studies on how the pubic hair/area microbiome behaves under controlled circumstances should be done to set up some models to provide prior odds of a sample being a mixture for Bayesian models. As shown in Chapter 4, *in silico* modeling of mixtures provided a proof of concept that using percent OTUs shared between an evidence sample and a known reference from a suspect could be sensitive and specific when the suspect contributed at least 25% and the percent OTU sharing was above 55%. A model could be constructed using the percent OTU sharing to give a likelihood of having the microbiome profile if the suspect

has a contribution of at least 25% compared to an unknown contributor. Because it is unknown how much of an individual's microbiome is transferred during sexual activity, controlled studies tracking the microbiome before, immediately after, and for set time points following sexual activity, with or without a marker species such as used by Kort et al. [11], would add to the developing model.

Temporal intra-individual variation also plays a role in how well a model of microbiome mixing performs. While SourceTracker, a Bayesian tool for predicting the proportional contribution of sources to a sample [12], performed well in predicting proportions of contributors to an *in silico* model, its success with collected population samples with known sexual activity but unknown mixture ratios was limited. While the inter-individual variation was greater than the intra-individual variation so that error rate Random Forest predictions of identity was low (10%), the SourceTracker analyses seemed to be affected by the intra-individual variation. In the case of the Random Forest models, subsets of the data were used to set up and train the model over 500 iterations. By doing these iterations, the intra-individual variation is accounted for and used in the model. SourceTracker was originally used with environmental samples where the question was which potential sources, to include oral, gut, skin, and soil, contributed to an environmental sample. In these cases, multiple samples contributed to each source, much in the same way multiple samples in a subset were used to make the Random Forest model. In testing SourceTracker with the pubic hair/pubic swab samples, only one sample per individual was used as a

source sample. Further work could be done to show what the effect of using multiple samples from an individual collected over time as a source would have on the SourceTracker analyses.

An additional adjustment to the priors for Bayesian determinations of the presence of a mixture may include prior likelihood of sexual activity. The Spring 2017 American College Health Association-National College Health Assessment (ACHA-NCHA) [13] reports 31.2% (32% of males and 31% of females) of participants (total n=63,497) with no sexual partners, defined as partners with whom they had oral sex, vaginal intercourse or anal intercourse, within the last 12 months. Among those who were sexually active, the mean number of sexual partners was 1.62 (1.90 for males, 1.46 for females). For vaginal intercourse, 32.9% of reported never having performed this sexual activity, 19.3% reported having done this sexual activity but not in the last 30 days, and 47.8% reported having done this sexual activity in the last 30 days. An *a priori* probability of a sample being a mixture can be provided by using this sexual activity data.

All of these models, whether they be Random Forest or Bayesian, must first be framed by the appropriate forensic question. First, it must be determined whether there is a mixture or not. Then the likelihood of a microbiome given various conditions can be considered. Potential questions include: What is the probability of a given microbiome profile from a recovered pubic hair if an individual is the source of the evidence? What is the probability of a given a microbiome profile

from a victim's pubic hair sample/pubic swab if the victim and a suspect are contributors to the microbiome profile? What is the probability of a given microbiome profile from a victim's pubic hair sample/pubic swab if the victim and a suspect from a pool of suspects are contributors to the mixture. For a pool of suspects, each suspect may be evaluated separately and the microbiome profile more or less likely for each suspect, eliminating some from being a contributor.

In the case of sexual assault, excluding an individual from a given mixture can be just as powerful as including an individual. In terms of inclusion/exclusion of an individual to a presumed single-source evidence sample, Random Forests modeling performs well given multiple reference samples from an individual to account for sampling and temporal variation. Blind retesting of hair samples resulted in 9 out of 10 correct re-identifications using the Random Forest model. This model also provides information on how often in 500 model trees and individual is predicted, adding weight to the predictions. For questions of presence of an individual in a mixture, percent OTUs shared and SourceTracker show promise with further studies to refine the models. Just as with Random Forest, multiple samples from an individual to form a reference may improve outcomes. Samples from a larger pool of individuals would also help to refine upper and lower percent OTU sharing within and between individuals in order to establish cut-offs to exclude an individual from a sample. Overall, preliminary results here show great potential for the development of models for answering forensically-

relevant questions using the pubic hair/pubic area microbiome in sexual assault cases.

5.5 Additional Considerations

Given that forensic microbiome analysis would need to be conducted in financially-limited public forensic laboratories, all steps in the process need to be evaluated for cost and time effectiveness versus return of information or helpfulness. In the pubic hair/pubic area microbiome, the concentration of the DNA extracts consistently fell below the limits of detection of the fluorometer, even after concentration of the extracts with a vacuum concentrator. Given that the optimum input DNA amount for the targeted amplification of the 16S rRNA gene is in the dynamic range of the fluorometer, it was known that the extracts contained less than that amount. While exact quantification could have been performed using qPCR at this step, further quantification was not performed in the interest of time and money. The use of qPCR was most beneficial post-PCR for the purposes of normalizing the libraries. A cost-benefit analysis would also need to be considered in how many samples to multiplex in a single library. The MiSeq reagent cartridge used here is capable of handling up to 96 samples in a library. For this study, 40-50 samples were combined in a single library. As more samples are included in the library, the cost-per-sample for analysis is reduced. Other sequencing technologies, such as the ThermoFisher Ion GeneStudio S5 which

can handle fewer samples in a sequencing run at lower cost, may be investigated. However, this platform has not been as thoroughly used as the MiSeq with microbiome samples. At what threshold this cost-per-sample becomes acceptable and how many samples the crime laboratory expects to need analyzed over a given time frame will need to be balanced. For forensic laboratories that will be using sequencing technology for human DNA analysis, the additional expense of implementing microbiome analysis should not be as great.

5.6 Conclusion

The limitations and concerns addressed here are not insurmountable ones. Further studies confirming the individuality of the pubic hair/pubic microbiome and refinement of detection of mixtures will move this methodology forward towards use in the crime laboratory. It should be possible to develop Bayesian methods for the weighting of the significance of matches. It would be extremely useful to conduct mock analyses utilizing archived case samples, but ethical and legal barriers will need to be overcome. Nevertheless, overall, the pubic hair/pubic area microbiome shows great potential for addition to the crime laboratory toolkit.

5.2 References

[1] Federal Bureau of Investigation. Quality Assurance Standards for Forensic DNA Testing Laboratories. effective date September 1, 2011 ed.

[2] Scientific Working Group on DNA Analysis Methods (SWGDAM) Committees.

[3] Clarke TH, Gomez A, Singh H, Nelson KE, Brinkac LM. Integrating the microbiome as a resource in the forensics toolkit. Forensic Science International: Genetics. 2017;30:141-7.

[4] Budowle B, Connel ND, Bielecka-Oder A, Colwell RR, Corbett CR, Fletcher J, et al. Validation of high throughput sequencing and microbial forensics applications. Investigative Genetics. 2014;5:9.

[5] Lim MY, Song E-J, Kim SH, Lee J, Nam Y-D. Comparison of DNA extraction methods for human gut microbial community profiling. Systematic and Applied Microbiology. 2017.

[6] de Boer R, Peters R, Gierveld S, Schuurman T, Kooistra-Smid M, Savelkoul P. Improved detection of microbial DNA after bead-beating before DNA isolation. Journal of Microbiological Methods. 2010;80:209-11.

[7] Jones MB, Highlander SK, Anderson EL, Li W, Dayrit M, Klitgord N, et al. Library preparation methodology can influence genomic and function predictions in human microbiome research. PNAS. 2015.

[8] Illumina. 16S Metagenomic Sequencing Library Preparation. Part # 15044223 Rev. B ed.

[9] Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, et al. A large genome center's improvements to the Illumina sequencing system. Nat Methods. 2008;5:1005-11.

[10] Bright J-A, Taylor D, McGovern C, Cooper S, Russell L, Abarno D, et al. Developmental validation of STRmix TM, expert software for the interpretation of forensic DNA profiles. Forensic Science International: Genetics. 2016;23:226-39.

[11] Kort R, Caspers M, van de Graaf A, van Egmond W, Keijser B, Roeselers G. Shaping the oral microbiota through intimate kissing. Microbiome. 2014;2:41.

[12] Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG, et al. Bayesian community-wide culture-independent microbial source tracking. Nature Methods. 2011;8:761-5.

[13] Association ACH. American College Health Association-National College Health Assessment (ACHA-HCHA-II) Reference Group Data Report - Spring 2017. 2017.